ALTERNATIVE NUCLEIC ACID:
THE EXPANDED GENETIC ALPHABET

By

STEFAN LUTZ

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE
UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1999

Dedicated to my family

for their endless love and support

ACKNOWLEDGMENTS

First and foremost, I would like to thank Dr. Steven A. Benner for the opportunity to work in his lab. Without his support and enthusiasm, this work would not have been possible.

My thanks go to the past and present members of the nonstandard Benner research group, including Dr. Petra Burgstaller, Dr. Tom Battersby, Dr. Kevin Devine, Heike Held, Dr. Michael Lutz, Darwin Ang, Mike Thompson, Mat Carrigan, Vic Ohtamaa, Dr. Janos Kodra, Dr. Jeong Ho Park, Mike Sismour, Dr. Simona Jurczyk, and my successor, Cynthia Whitehead. A special thanks goes to my two undergraduate researchers, Ken Park and Kim Ferguson, for all their input into this research.

For the supply of precursor material, I would like to acknowledge Andre Mueller (ETH, Zurich; 2,4-diamino5-ribofuranosyl-pyrimidine) and Dr. Simona Jurczyk (Sulfonics, Eragen, Alachua FL) for the 2'-deoxyxanthosine phosphoramidite. For experiments involving the triphosphate of pyDAD, the material prepared by Jen Horlacher was greatly appreciated. Dr. Jeong Ho Park and Mike Sismour kindly provided the HIV type-1 reverse transcriptase for the incorporation studies discussed in Chapter 3.

I am also very grateful to Yves Choffat (Zoologisches Institut, Univeristy of Zurich), Jace Dinehart (ICBR DNA synthesis core facility, University of Florida), and Dr. Tom Battersby for the extra effort in synthesizing the oligonucleotides; containing the nonstandard nucleotides.

Furthermore, I wish to thank our collaborators: Dr. G. Folkers (ETH Zurich; phosphorylation of nonstandard nucleotides by Herpes Simplex Virus type-1 thymidine kinase), Dr. Mary Kopka (Dickerson Laboratory, UCLA; crystallization of oligonucleotides containing nonstandard basepairs), and the folks at New England Biolabs (Beverly, MA; support in cloning and expression of thermostable DNA polymerases).

Last, but not least, I want express my thanks to all the poor fellows who actually volunteered to read and correct this work at one stage or another: Dr. Petra Burgstaller, Dr. Kevin G. Devine, Dr. Tom Battersby, Dr. David Liberles, Lisa Manning, and Jennifer Cottone.

TABLE OF CONTENTS

SYMBOLS AND ABREVIATIONS

A	adenosine
AIBN	2,2'-azobisisobutyronitrile
AIDS	acquired immunodeficiency syndrome
Amp	ampicillin
AMV	avian myeloblastosis virus
arom	aromatic
APS	ammonium persulfate
ATP	adenosine triphosphate
Bn	benzyl
bp	base pair(s)
BSA	bis(trimethylsilyl)acetamide
C	cytidine
Ci	Curie (1 Ci = 3.7×10^7 Bequerel)
CM	chloramphenicol
COSY	2-dimensional correlation spectroscopy
CPG	controlled pore glass
cpm	counts per minute
d	days
Da	Dalton
ddN	dideoxynucleoside (ddA, ddC, ddG, ddT)

dN	deoxynucleoside (dA, dG, dC, ψT, dψU, d(pyDAD), d(puADA))
DBU	1.8-diazabicyclo[5.4.0]undec-7-ene
DMAP	4-dimethyl aminopyridine
DMF	N,N-dimethyl formamide
DMSO	dimethyl sulfoxide
DMT	4',4''-dimethoxy trityl
DTT	1,4-dithio-DL-threitol
ε	absorption coefficient ($M^{-1}cm^{-1}$)
E. coli	*Escherichia coli*
EDTA	ethylendiamino tetraacetate
ESI	electrospray ionization
EtOH	ethanol
FAB	fast atom bombardment
G	guanosine
h	hours
HIV-1	human immunodeficiency virus type-1
HMDS	hexamethylendisilazane
HPLC	high performance liquid chromatography
HV	high vacuum
IPTG	isopropyl-β-D-thiogalactopyranoside
IRR	irritation radiation (NOE)
Kan	kanamycin
λ	wave length (nm)
LB	Luria-Bertani medium

LB-Agar	Luria-Bertani medium (w/ Agar)
LSC	liquid scintillation counting
MALDI TOF	matrix assisted laser desorption ionization time of flight
MeOH	methanol
MHz	mega hertz
M-MuLV	moloney murine leukemia virus
min	minutes
MS	mass spectroscopy
MTP	microtiterplate
NPA	m-nitrophenylalcohol
NOE	nuclear Overhauser effect
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
Pfu	*Pyrococcus furiosus*
P/T	primer-template complex
ppm	parts per million
Pwo	*Pyrococcus woesi*
puADA	xanthosine (X)
pyDAD	2,4-diamino-2'-deoxy-(5-β-D-ribofuranosyl) pyrimidine
RP	reverse phase
RT	room temperature
SDS	sodium dodecylsulfate
SPA	scintillation proximity assay

T	thymidine
Taq	*Thermus aquaticus*
TFMSA	trifluoromethanesulfonic anhydride
ψT	pseudothymidine
ψU	pseudouridine
TBAF	tetrabutylammonium fluoride
TBE	Tris / borate / EDTA buffer
TE	Tris / EDTA buffer
TEA	triethylamine
TEAA	triethylammonium acetate
TEMED	N,N,N,N-tetramethylethylen diamine
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
TMS	trimethylsilyl
Tris	tris(hydroxymethyl)aminomethane
Tth	*Thermus thermophilus*
wt	wild-type
U	uracil
X	xanthosine (puADA)

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ALTERNATIVE NUCLEIC ACID: THE EXPANDED GENETIC ALPHABET

By

Stefan Lutz

August 1999

Chairman: Dr. Steven A. Benner
Major Department: Chemistry

Nucleic acid can perform multiple functions. Beside storing and transferring genetic information, both RNA and DNA can fold into tertiary structures, converting them into receptors, ligands, and catalysts. Nevertheless, having only four nucleic acid building blocks, DNA has limited functionality and versatility.

Following the Watson-Crick rules for base pairing, additional, nonstandard base pairs can be designed, expanding the genetic alphabet in principle to 12 letters. The physicochemical and biological properties of one of these nonstandard base pairs, between 2,4-diamino-2'-deoxyribofuranosyl-pyrimidine (pyDAD) and xanthosine (puADA), was studied in this thesis.

First, the two nonstandard nucleotides were synthesized and incorporated into oligodeoxyribonucleotides by automated DNA synthesis. A series of melting studies on a dodecamer that contained two pyDAD/puADA base pairs were performed. These confirmed the formation of stabilizing hydrogen bonds between the nonstandard nucleotides.

Second, the enzymatic incorporation of pyDAD and puADA by HIV type-1 reverse transcriptase was investigated. The enzyme showed specific and template-directed incorporation of the two nonstandard nucleotides opposite each other. Strong pausing after incorporation of the nonstandard base pair made the amplification of sequences with multiple nonstandard bases impossible, however.

Third, the effect of C-nucleosides on a polymerase's performance was studied. The strong pausing of polymerases at positions of the nonstandard base pair was partly attributed to structural disturbances of the primer-template complex caused by the C-nucleosidic pyDAD. Using pseudothymidine, a C-glycoside having the same standard hydrogen bonding pattern of thymidine, this hypothesis was tested. In a newly developed high-throughput assay based on primer extension and scintillation proximity counting, *Taq* DNA polymerase was found to incorporate several pseudothymidine nucleotides sequence-specifically over ten rounds of PCR-like amplification. Kinetic data indicated only a minor effect on the accuracy of primer extension.

Finally, methodology was estabilished to perform directed molecular evolution on thermostable DNA polymerases. Applying random mutagenesis and DNA shuffling to the gene of 9°N DNA polymerase, followed by ligation and transformation into an expression host, thousands of low-frequency mutants of the DNA polymerase were generated. The performance of the enzyme library was studied by high-throughput screening in primer extension reactions and scintillation proximity assays.

CHAPTER 1
INTRODUCTION

Nucleic acid is one of the key elements of life. Made of four building blocks, adenosine, guanosine, cytidine, and thymidine/uridine, nucleic acid functions as a storage and transfer system for genetic information. According to the revised central dogma in molecular biology (Figure 1-1), genetic information stored as DNA can be replicated directly by DNA polymerases. In cases where RNA serves as the storage device, the information is first transcribed back into DNA by reverse transcriptases, followed by regular replication, and finally transcription back into RNA. Transcription is catalyzed by RNA polymerases. Finally, RNA or more specifically messenger RNA (mRNA) then is translated into proteins by ribosomes, a hybrid of protein and ribosomal RNA (rRNA). The various amino acid building

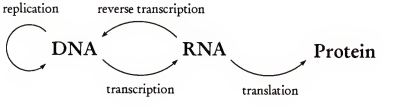


Figure 1-1. The revised central dogma of molecular biology. DNA and RNA both function as genetic storage systems. RNA also performs in multiple roles during protein synthesis. Besides acting as the information transfer system (messenger RNA), it carries amino acid building blocks to the site of synthesis (transfer RNA), and finally forms an essential part of the actual catalyst for protein synthesis (ribosomal RNA). Proteins in turn form the catalysts that allow DNA replication and transcription, as well as conversion of RNA into DNA by reverse transcription.

blocks are delivered by the transfer RNA (tRNA), shuttling between protegenic tRNA synthetases, that load the amino acid onto the tRNA, and the ribosomes.

In evolutionary terms, this amazing machinery holds one of the biggest mysteries – the "chicken and egg" dilemma of how to generate proteins without nucleic acid and how to replicate nucleic acid without proteins. As a possible solution, the hypothesis of the RNA-world was proposed about 30 years ago (Crick, 1968; Orgel, 1968; Woese, 1967). It suggested that a single type of molecule, RNA, was responsible for both genetics and catalysis. The proposal had a major weakness, however, since RNA catalysis was unknown.

Ribozymes and Deoxyribozymes – Novel Catalysts from Nucleic Acid

Nucleic acid has long been viewed as a chemically "inert" macromolecule, designed for the storage of genetic information and not suitable for catalysis. This concept was first challenged by a report of RNA's specific and autocatalytic self-cleavage in the presence of Pb^{2+} (Werner et al., 1976). Almost one decade later, Cech and coworkers (Kurger et al., 1982) observed the self-excision of an intervening sequence (IVS) in isolated preribosomal RNA from *Tetrahymena*. The processing of the rRNA involved strand cleavage to remove the IVS, followed by religation of the newly generated rRNA ends and was found to be completely protein-independent. Due to its enzyme-like characteristics, the term ribozyme was introduced to describe "RNA molecules with the intrinsic ability to break and form covalent bonds" (Kurger et al., 1982).

Shortly after, Altman and coworkers (Guerrier-Takada et al., 1983) reported the first truly catalytic ribozyme. They found that at high magnesium concentrations, the RNA subunit of RNase P alone was sufficient to carry out the catalytic cleavage of tRNA

precursors. For their discoveries of the catalytic activity of RNA, T. R. Cech and S. Altman shared the 1989 Nobel Prize in chemistry.

These discoveries removed the main objection to the RNA world. Nucleic acid itself could act as a catalyst. It was therefore no longer unthinkable that RNA, at one point in time, was the one and only molecule to store the genetic information in an organism, as well as catalyze the chemistry to keep the system "alive". Although it is unclear today where the larger RNA fragments came from originally, a critical gap in the hypothesis is the failure to produce a self-replicating ribozyme, the holy grail of the RNA world and the true key to life.

But ribozymes not only open the door for answers to question from the past. RNA and DNA-based catalysts represent an interesting concept for the next generation of drug design. Phosphodiester bond cleavage for example, the chemistry that ribozymes do best, can be used to control gene expression in vivo (Seyhan et al., 1998; Sioud, 1999). In combination with easy controllable sequence specificity, based on the length of the hybridization sequence, messenger RNA in cancer cells or from viral infections such as Hepatitis or Herpes, as well as the genomic RNA from HIV can in principle be targeted and hydrolyzed in the cell.

In summary, the prospect of understanding how life originated, as well as the potential for new therapeutic and diagnostic applications of ribozymes have driven research. Mechanistic and structural studies on one hand and improved methodology on the other have helped develop quicker ways to identify new activities, and search for ways to further improve the performance of nucleic acid catalysts.

In Vitro Selection/SELEX

An in vitro technique for searching and identifying RNA-based receptors or catalysts was introduced by Joyce (Joyce, 1989). In his paper, he proposed the combination of PCR

amplification, mutation, and selection to isolate RNA strands with desired properties from a pool of random oligoribonucleotides. Used in an iterative fashion, the method resulted in the enrichment of sequences that performed best under the applied selection pressure. Called in vitro selection or SELEX (systematic evolution of ligands by exponential enrichment), this technique allowed the exploration of the potential of nucleic acid for selective binding and catalysis in a simple laboratory system. In addition, the method did not require any in vivo steps, drastically speeding up the procedure. The idea was tested shortly after. In vitro selection and SELEX were applied successfully to isolate sequences of the *Tetrahymena* ribozyme with improved catalytic performance (Robertson & Joyce, 1990), as well as to identify RNA ligands with a high binding affinity for dyes (Ellington & Szostak, 1990) and a protein (Tuerk & Gold, 1990).

In contrast to Joyce's originally proposed method of amplification, mutation, and selection, SELEX or in vitro selection does not include a intentional random mutagenesis step as part of each round of selection. The final target sequence therefore had to be part of the initial RNA or DNA oligonucleotide library. For a 100-base random oligonucleotide, a typical library, containing approximately 10^{15} members, samples only a tiny fraction of all possible combinations ($4^{100} = 10^{60}$). The chances to even have the "best" sequence present in the initial pool is extremely small, not considering the odds of losing the sequence by accident over the multiple rounds of selection. The inflexibility of the in vitro selection system (once lost, forever lost) can be overcome by in vitro evolution (Beaudry & Joyce, 1992).

In Vitro Evolution of Nucleic Acid

In vitro evolution adopts the same basic principle as in vitro selection but adds the element of low-level random mutagenesis to the system. This additional element allows true

Darwinian evolution: i) introduction of genetic variation, ii) selection of individuals based on some fitness criteria, and iii) reproduction of selected individuals. Multiple rounds of in vitro evolution hypothetically allow the exploration of the entire sequence space of a given oligonucleotide sequence. An initial pool of random oligonucleotides does not necessarily have to contain the "best" sequences as in the SELEX or in vitro selection experiments. Instead, each selection cycle identifies and isolates the top performers of a given pool which are then diversified by random mutagenesis, creating a new pool which carries the beneficial mutations from the previous rounds.

In summary, in vitro selection and evolution of RNA and more recently also DNA oligonucleotides (for review see Breaker, 1997a; Breaker, 1997b) have resulted in a variety of enzymatic activities (Table 1-1). While natural ribozymes identified so far perform only simple tasks such as phosphoester transfer or hydrolysis, in vitro techniques have expanded the repertoire of RNA and DNA catalyzed reactions significantly. The ability of nucleic acid to perform C-N and C-C bond formation represent two examples.

Although the search for novel catalysts is far from being exhaustive, to obtain more sophisticated nucleic acid enzymes such as a self-replicating ribozyme, the "holy grail" of the field, may require an extra effort. On one side, Bartel and coworkers (Sabeti et al., 1997) suggest that to find rare or more complex functionality in ribozymes, the average oligonucleotide length in a library would have to be increased from typically 70 positions to 200 bases and more. They argue that the longer sequence increases the probability of modular organization of smaller motifs within the oligonucleotide and thereby raising the chances for detection of rare activities. On the other hand, an increase in diversity and functionality of the nucleotides in an oligomer, similar to amino acids in proteins, could move nucleic acid enzymes to the next level of catalysis.

Table 1-1. RNA and DNA enzyme catalyzed reactions (based on Joyce, 1998)

Reaction	RNA enzyme	DNA enzyme	Ref.
Phosphoester transfer	X*		(Beaudry & Joyce, 1992)
Phosphoester hydrolysis	X*⁺	X	(Breaker & Joyce, 1994; Carmi et al., 1998; Geyer & Sen, 1997; Robertson & Joyce, 1990)
Polynucleotide ligation	X	X	(Cuenoud & Szostak, 1995; Hager & Szostak, 1997; James & Ellington, 1997)
Polynucleotide phosphorylation	X	X	(Li & Breaker, 1999)
Mononucleotide polymerization	X		(Been & Cech, 1988)
Aminoacyl transfer	X		(Illangaskare & Yarus, 1997; Lohse & Szostak, 1996)
Amide bond cleavage	X		(Dai et al., 1995)
Amide bond formation	X⁺		(Wiegand et al., 1997)
Peptide bond formation	(X*)		(Zhang & Cech, 1997)
N-alkylation, *S*-alkylation	X		(Wilson & Szostak, 1995)
Porphyrin metallation	X⁺	X⁺	(Li & Sen, 1996)
Diels-Alder cycloaddition	X⁺		(Seelig & Jaeschke, 1999; Tarasow et al., 1997)
Isomerase activity	X		(Prudent et al., 1995)
Nucleotide synthesis	X		(Unrau & Bartel, 1998)
Oxidative DNA cleavage		X	(Carmi et al., 1996)

List of reactions, catalyzed by RNA or DNA enzymes. RNA catalytic activities, found in nature are indicated with X*; all other catalysts were isolated from in vitro experiments. Nucleic acid enzymes that utilize metal ions other than magnesium or functionalized nucleotides are marked X⁺.

The Secrets of Nucleic Acid Catalysis

The key for the successful catalysis by ribozymes lies within its structural abilities. For a long time, nucleic acid was viewed as a double helical "stick", held straight by the charge

repulsion of the phosphate backbone. It was thought very unlikely that DNA and RNA could fold into a defined three-dimensional structure, a prerequisite for successful catalysis. X-ray crystallographic structures of tRNA told a different story, however (Kim et al., 1973). In particular, single-stranded RNA was shown to adopt a wide diversity of tertiary structures, creating cavities that could accommodate binding sites for substrates and act as an active site for catalysis. Today, further evidence for the structural capabilities of RNA and DNA aptamers is provided by binding constants in the nanomolar range reported for aptamer-substrate complexes (Breaker, 1997a; Joyce, 1994; Lorsch & Szostak, 1994; Tuerk & Gold, 1990).

In comparison to proteins, however, ribozymes perform rather poorly as catalysts. Most rate constants for the reactions, listed in Table 1-1, are three-to-five orders of magnitude slower than their protegenic counterpart. In addition, catalytic ribozymes identified so far require some kind of cofactor, mostly metal ions. These observations suggest that standard nucleic acid lacks the functional diversity found in proteins to perform as an efficient catalyst.

One possible solution to this problem is derivatization of existing nucleotides as is well documented for tRNA.

Structurally and functionally modified nucleotides in nature. Nature frequently varies functionality to manipulate the physical properties of nucleic acid. Approximately sixty naturally modified nucleosides, most of them found in tRNA, have been reported so far (Bjoerk, 1995; Limbach et al., 1994). All except one of these nucleotides are generated by posttranscriptional modification of standard nucleotides. The most common substitution is methylation on the nucleobase and/or deoxyribose sugar. In addition, nucleotides are derivatized with amino acids such as lysine into lysidine (Muramatsu et al., 1988b) and

converted into C-nucleosides such as pseudouridine and pseudothymidine (Argoudelis & Mizsak, 1976).

The importance of modified ribonucleotides in Nature is underlined by two observations. First, the expenditure of sequence space to maintain the enzyme machinery, required for the postsynthetic modifications, consumes approximately 1% of *E. coli*'s genome (Bjoerk, 1995). Second, the replacement of modified nucleotides in tRNA with standard nucleotides (for example pseudouridine with uridine) caused a drastic reduction in translation efficiency (Bjoerk, 1995). Mechanistically, little is known. Modifications located directly in the anticodon region alter the recognition sequence of tRNA (Muramatsu et al., 1988a; Tomita et al., 1999), while their presence in other regions of the molecule seems to contribute to the conformational rigidity or flexibility of particular sub-structures (Bjoerk, 1995; Davis, 1995).

The presence of these modified nucleotides in all three kingdoms (prokaryotes, eukaryotes, and archae) suggests that they were present in the most recent common ancestor. This suggests that they were significant in the early life on earth. On the other hand, the fact that most of today's modifications are introduced posttranscriptionally by enzymes represents a clear weakness in the hypothesis and awaits further studies.

The Next Generation Nucleic Acid Enzymes

Today's DNA and RNA building blocks are presumably the result of evolution of DNA to be a reliable genetic macromolecule, not a catalyst. Its physical and chemical properties have been adapted to optimal performance in tasks such as replication and transcription. Being comprised of only four structural and chemically very similar building blocks, nucleic acid does not possess functionality normally important in protein catalysis. Neither a general acid/base such as histidine nor charged residues like lysine, arginine, or

carboxylic acids exists in nucleic acid. Furthermore, sulfhydryl groups as in cysteine that could act as nucleophiles are not found in DNA and RNA.

Although some functionality can be partially substituted by various regions in the nucleotides such as the 2'-hydroxy group in RNA and the phosphate backbone, the functional limitation has been blamed for the restrained repertoire of reactions that are catalyzed by nucleic acid (Table 1-1). An increase in functionality and/or diversity could expand the scope of RNA and DNA catalysis, as well as improve the performance of existing catalysts. The literature describes four major approaches to raise the structural and chemical diversity of nucleic acid.

Metal ion cofactors. The variation of metal ions in ribozymes can change the chemical and physical properties of an oligonucleotide. Most nucleic acid enzymes rely on a metal cofactor, commonly magnesium. Metal ions can participate in catalysis by acting as a Lewis acid or through coordinated water, which are deprotonated and then act as a Lewis base (Steitz & Steitz, 1993). Furthermore, they play a critical role in stabilizing nucleic acid tertiary structures (Hampel & Cowan, 1997). The substitution of magnesium ions with calcium (Faulhammer & Famulok, 1996), transition elements such as manganese, copper, and lead (Chartrand et al., 1997; Peracchi et al., 1997), or lanthanum (Tsubouchi & Bruice, 1995), affected the catalytic activity and substrate specificity of the ribozyme. The variation of the metal ions in a selection experiment is therefore likely to produce novel reactivity and structural features in oligonucleotides, which can raise the catalytic performance of ribozymes.

Functionalized nucleic acid. A second approach to increase the catalytic potential of nucleic acid is by replacing standard nucleotides with nucleotide analogues that have enhanced functionality as shown in Figure 1-2. In recent years, the chemical synthesis of a series of such 5-position functionalized uridines has been reported (Cook et al., 1988; Dewey et al., 1996;

Kahl & Greenberg, 1999; Latham et al., 1994; Sakthivel & Barbas, 1998). The two main reasons for choosing uridine as a starting structure for these experiments are based on the easy accessibility of modified nucleotides from the commercially available 5-iodouridine, as well as the possible role of uridine and its substituted analogues in prebiotic chemistry (Robertson & Miller, 1995).

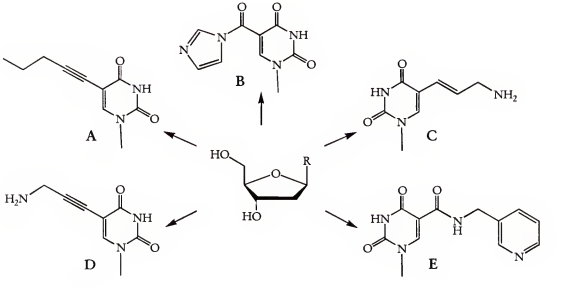


Figure 1-2. Functionalized ribo- and 2'-deoxyribonucleosides: A) 5-(1-pentynyl)-2'-deoxyuridine (Latham et al., 1994), B) 5-imidazole-2'-deoxyuridine (Wiegand et al., 1997), C) 5-(3-aminopropenyl)-2'-deoxyuridine (Cook et al., 1988; Sakthivel & Barbas, 1998), D) 5-(3-aminopropynyl)-2'-deoxyuridine (Battersby et al., 1999), E) 5-pyridylmethylcarboxamid-uridine (Tarasow et al., 1997).

Using a simple olefin side chain in 5-(1-pentynyl)-2'-deoxyuridine (Figure 1-2A), Latham and coworkers (Latham et al., 1994) performed in vitro selection for a thrombin-binding aptamer. The isolated pool of binding oligonucleotides had a completely different sequence composition then the pool obtained in similar experiments with only standard nucleotides, indicating the selection of novel sequence motifs due to the functionalized base.

Wiegand and coworkers (Wiegand et al., 1997) demonstrated in their selection for an RNA amide synthase that nucleotide analogues can improve reaction rates. Replacing uridine with its 5-imidazole analogue (Figure 1-2B), the authors isolated an RNA enzyme with an approximate rate enhancement of five orders of magnitude over background. Unfortunately, a direct comparison between the catalytic performance of the sequence libraries with and without the modified nucleotides is not possible. Although the authors mentioned a "significantly reduced rate enhancement" in the absence of the 5-imidazole analogue, no kinetic data were reported. Finally Tarasow and coworkers (Tarasow et al., 1997) developed the first ribozyme that catalyzes the formation of a carbon-carbon bond. The substitution of 5-pyridylmethyl-carboxamide-UTP (Figure 1-2E) for UTP lead to the isolation of a Diels-Alder catalyst, a mechanistic task not accomplishable by standard nucleotides in selection experiments.

In summary, these three examples support the hypothesis that functionalized nucleotide analogues improve the catalytic properties of nucleic acid. However, this approach faces important restrictions. First, the nucleotide analogue must maintain the genetic capabilities of nucleic acid. The substituted uridine triphosphate must be utilized by polymerases and, in turn, must also be able to serve as a template during replication. Second, the nucleotide analogue can not be substituted partially; a mixture of uridine and uridine analogues can not be faithfully replicated since the Watson-Crick basepairs for uridine and its substitute are the same. Therefore, the maximum number of functionalized nucleotides in nucleic acid is limited to four building blocks. One way to overcome the limitation of having only four building blocks in nucleic acid is to add cofactors that are not covalently bound to the nucleotides.

Non-covalent cofactors molecules. The presence of small organic molecules that are non-covalently bound to the ribozyme should in principle improve the catalytic performance of nucleic acid. Separate cofactor molecules in the reaction mixture have far-reaching potential. No longer limited to chemical and structural diversity of nucleic acid, new catalysts could be isolated from RNA or DNA pools in the presence of libraries of small organic molecules. Only recently, Roth and Breaker (Roth & Breaker, 1998) have reported for the first time the successful in vitro selection of a histidine-dependent deoxyribozyme for RNA cleavage. The sequence does not require any metal ions for catalysis but presumably cleaves RNA by a mechanism similar to that used by ribonuclease A, directly involving the histidine cofactor.

One of the problems associated with cofactors in solution is the competition between tight cofactor binding and catalytic efficiency. A survey of Roth's sequence pool after several rounds of selection showed an inverse correlation between increased histidine-binding affinity and catalytic rate enhancement. Catalysis obviously requires not only tight binding but also proper orientation of the cofactor with respect to the substrate for optimal activity. In conclusion, the co-optimization of the binding affinity and catalytic activity of a bimolecular system has proven difficult but possible.

To develop even better nucleic acid-based catalysts, researchers now try to combine the effectiveness of functionalized nucleotides and the almost unlimited diversity of non-covalent cofactor molecules. By increasing the number of nucleotides from four to six or more building blocks, the complexity of an oligonucleotide raises, increasing the chances to observe protein-like performance. Additional nucleotides that form a basepair, distinguishable from adenosine-thymidine and guanosine-cytidine have been reported previously (Piccirilli et al., 1990; Strazewski & Tamm, 1990; Switzer et al., 1989), expanding the genetic alphabet from four up to twelve letters.

Expanding the Genetic Alphabet

Watson and Crick, as part of their work on the X-ray diffraction of DNA (Watson & Crick, 1953), defined the two rules of complementary for base pair formation in nucleic acid. According to these rules, formation of a correct base pair is based on size complementary (large purine pairs with small pyrimidine) and hydrogen bonding complementary (hydrogen donor from one nucleobase pairs with an acceptor from the other). In Nature, the adenosine-thymidine/uridine and the guanosine-cytidine base pairs represent these rules (Figure 1-3). However, with three hydrogen bonds forming a base pair, the number of possible combinations can be extended to a total of eight hydrogen bonding patterns. With six of the eight possible base pairs chemically feasible, the genetic alphabet could be expanded from four to twelve letters (Figure 1-3).

The pyAAD/puDDA basepair. The pyAAD/puDDA basepair, also known as isocytidine (isoC) and isoguanosine (isoG), was the first nonstandard base pair to be discussed (Rich, 1962). This base pair may be the closest structural analog of the standard base pairs. Like A, T, G, and C, isoC and isoG are both N-nucleosides. Further, the pyrimidine and purine skeletons are retained.

IsoC is more reactive than its natural isomer C. For example, isoC undergoes deamination with a half-life $t_{1/2}$ = 30 days at -20°C (Switzer et al., 1993) to form uridine. This is considerably faster than the deamination rate for cytidine derivatives, with deamination half-life's on the order of years. The stability of isoC can, however, be increased by methylation in 5 position (Tor & Dervan, 1993).

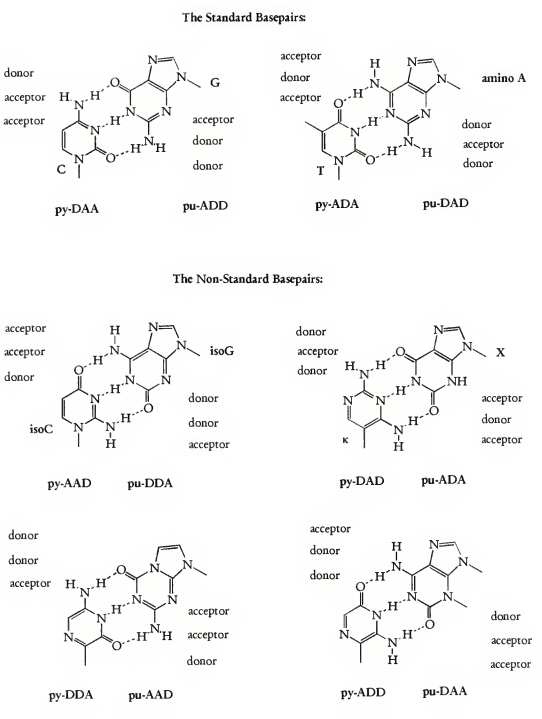


Figure 1-3. Summary of possible basepairs within the constraints of the Watson-Crick geometry. Pyrimidine base analogs are designated by "py", purines by "pu". Hydrogen acceptors (A) and donors (D) are indicated by uppercase letters, starting in the major groove.

One feature of isoG that distinguishes it from its natural isomer G is the presence of substantial amounts of two tautomeric forms at equilibrium in aqueous solution. The keto/enol ratio is 9:1; although this favors the desired hydrogen bonding pattern, complementary to isoC, the presence of the enol-form (complementary to T in a Watson-Crick geometry) may be responsible for the incorporation of significant amounts of thymidine opposite isoG in various polymerase-catalyzed elongation reactions (Kodra, 1998).

The pyADD/puDAA and pyDDA/puAAD basepair. In early studies of the pyADD/puDAA base pair in oligonucleotides, significant levels of epimerization of the pyrimidine analog were observed. The epimerization of the nucleoside is presumably caused by the protonation of the sugar ring oxygen followed by loss of a proton from the NH in the pyrazine ring system, and the cleavage of the C1-O4 sugar bond (Voegel & Benner, 1994). Similar epimerization was observed in the pyDDA/puAAD base (von Krosigk & Benner, 1995).

The pyDAD/puADA basepair. The pyDAD/puADA basepair is, next to isoC/isoG, the best studied nonstandard base pair of the expanded genetic alphabet. Xanthosine is a natural purine nucleoside with an acceptor-donor-acceptor hydrogen bonding pattern, and is the last intermediate in the biosynthesis of guanosine. Because of concerns about the stability (depurination) and the low pK_A of xanthosine, the C-nucleoside *N1*-methyloxoformycin B (designated π) (Piccirilli et al., 1990) and 2'-deoxy-7-deazaxanthosine (Lutz et al., 1996) were previously used as an alternative puADA nucleoside.

The preparation of the pyrimidine nucleoside with a donor-acceptor-donor is more troublesome. The N-glycosidic diaminopyrimidine nucleoside must carry a positive charge on one nitrogen if it is to have the correct tautomeric form. Further, the C-nucleosidic 2,6-diaminopyridine analog (Piccirilli et al., 1991a) was unstable, presumably to oxidation, and

decomposed within days. The third variant, the 2,4-diamino-5-ribofuranosyl-pyrimidine, is stable and has the desired DAD hydrogen bonding pattern. A series of enzymatic studies were performed on the pyDAD/puADA basepair, and the search for a polymerase is well documented in the literature (Angerer & Ankenbauer, 1995; Horlacher et al., 1995; Lutz et al., 1996; Piccirilli et al., 1990).

The idea behind the expanded genetic alphabet. The expanded genetic alphabet combines the simplicity of nucleic acid chemistry and the sophisticated functional repertoire of protein catalysis. By maintaining the nucleosidic character of the building blocks, all of the technological advantages of automated chemical DNA and RNA synthesis are applicable. Furthermore, fast, cost-effective, and simple enzymatic methods such as the PCR for sequence amplification and/or sequence analysis are not available for proteins. Finally, each member of the expanded genetic alphabet can be functionalized individually. When combined, the twelve nucleotides can cover almost all of the basic functionality important for efficient catalysis such as general acid/base, acidic and basic residues, and nucleophiles.

Research Objectives

The nonstandard base pair between 2,4-diamino-2'-deoxy-5-ribofuranosyl-pyrimidine (pyDAD) and 2'-dexoyxanthosine (puADA), its base pairing properties, as well as the template-directed incorporation of the two building blocks by a thermostable DNA polymerase, was the subject of this thesis. First, the ability of pyDAD and puADA to form a stable base pair in an oligonucleotide had to be confirmed. Second, the effect of C-nucleosides on the performance of DNA polymerases during primer extension was addressed. Finally, a new methodology to identify a DNA polymerase that would accept nonstandard base pairs was developed.

CHAPTER 2
THE CHEMISTRY OF ALTERNATIVE NUCLEIC ACID

Introduction

The additional building blocks of the expanded genetic alphabet all fit the Watson-Crick geometry and display the hydrogen bonding complementarity of standard Watson-Crick base pairs. These facts, however, do not guarantee that they will contribute as expected to the stability of a duplex or that they will be accepted as substrates by enzymes that handle natural DNA. A first step towards understanding and predicting the properties and potential of nonstandard base pairs therefore involves the studies of their behavior as a part of small oligonucleotides and as enzyme substrates. As part of this thesis, one of the four nonstandard base pairs of the expanded genetic alphabet, the pyDAD/puADA combination, was chosen for further investigations.

Both nucleosides are accessible by organic synthesis. Protocols for the preparation of the two key products, the phosphoramidite and the triphosphate, of puADA (Horlacher et al., 1995; Van Aerschot et al., 1989) and pyDAD (Chu et al., 1977; Piccirilli et al., 1990) have been reported previously. The phosphoramidite is used for the automated DNA synthesis, while the triphosphate represents the most common derivative of nucleosides that is used as substrates by enzymes.

Synthetic Approaches for pyDAD and puADA

The puADA nucleoside

The phosphoramidite of puADA was synthesized by Dr. S. Jurczyk from 2'-deoxyguanosine in 9 steps with an overall yield of 20% (S. Jurzcyk – personal communication). Following Van Aerschot's protocol (Van Aerschot et al., 1989), the key step in the synthesis was the deamination of the amino group in 2-position with in situ-generated nitrous acid. For DNA synthesis with the phosphoramidite, both carbonyls on the purine base were protected as their (2,4-dinitrophenyl)ethyl ether. The deoxyribonucleoside triphosphate of puADA was prepared in one step from 2'-deoxyguanosine triphosphate. Direct deamination with nitrous acid, followed by HPLC purification yielded the desired product in 40% yield. The same methodology was applied for the preparation of tritium-labelled material.

The pyDAD nucleoside

The ribonucleoside of the pyrimidine analog pyDAD was prepared following the procedure of Chu and coworkers (Chu et al., 1977) (Figure 2-1, step 1 through 5). As outlined in Figure 2-1, the nucleobase is constructed onto the ribose skeleton in multiple steps. Despite the length of the route and its chronically low yields, it is still the only synthesis that successfully leads to the 2,4-diaminopyrimidine ribonucleoside. Alternative syntheses such as the amination of pseudouridine (Townsend et al., 1978) and the palladium-catalyzed coupling of the heterocycle to a glycal (Farr et al., 1990) all failed to produce the 2,4-diaminopyrimidine nucleoside in detectable amounts (Dr. M. Lutz, Dr. S. Jurzcyk – personal communication).

The second half of the pyDAD synthesis (Figure 2-1, step 5 through 15) includes the deoxygenation by Barton chemistry (Barton & Subramanian, 1976), followed by the

preparation of either the phosphoramidite **13** or the triphosphate **14** of pyDAD, and was first published by Piccirilli et al. (Piccirilli et al., 1990). The yield for the thirteen-step synthesis from ribose to the phosphoramidite is between 1 and 1.5%.

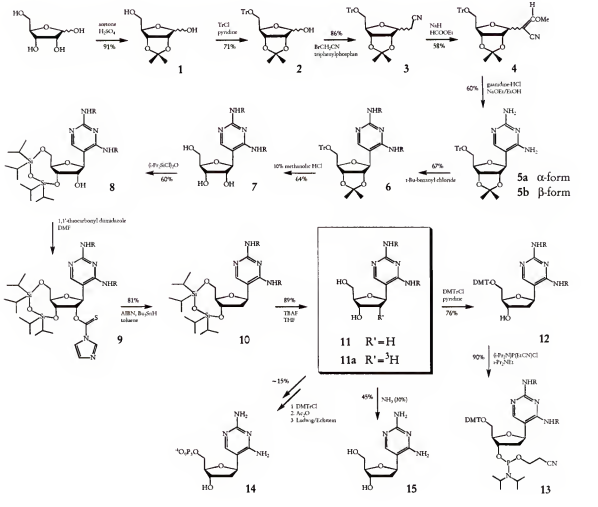


Figure 2-1. Synthetic route for the 2,4-diamino-5-(2'-deoxy-β-D-ribofuranosyl)-pyrimidine. Original (Chu et al., 1977; Piccirilli et al., 1990) and improved (Lutz & Benner, 1999) pathway are identical except for changes in reaction conditions and the protection group of the nucleobase which was modified from R= benzoyl to R= *tert* butyl benzoyl. The triphosphate **14** was prepared by phosphorylation of **11** (Horlacher, 1995).

Over the entire synthesis, two reactions stand out for their particularly low yields. The protection of the exocyclic amino groups on the pyrimidine (Figure 2-1, step 5-6) reportedly produced large amounts of side-products (Piccirilli, 1989), while the deprotection in the following step (Figure 2-1, step 6-7) resulted in a mixture of very hydrophilic compounds that was difficult to resolve. Over these two steps, the yield was only 20% and clearly needed to be optimized. Based on these results, a more hydrophobic protecting group such as *tert* butylbenzoyl instead of the commonly used benzoyl was suggested (Horlacher, 1995).

Structural features of the pyDAD nucleoside

Special attention was given to the configuration of the pyDAD deoxyribonucleoside. The formation of a base pair between pyDAD and puADA requires the correct stereochemistry at the anomeric centers (the β-form), as well as the anti-conformation of the nucleobases with respect to the sugar. Difference NOE spectroscopy is the fastest way to determine the absolute stereochemistry at the anomeric center and the preferred conformation of the N- or C-glycosidic bond (Davies, 1978; Knutsen et al., 1985). As shown in Figure 2-2, depending on the configuration at C-1', H-1' can either interact with H-3' (α-form) or H-4' (β-form). In addition, H-6 interacts with H-1' when the nucleobase is in the syn-conformation relative to the sugar, resulting in a change in signal intensity when irradiated.

Structural Effects of pyDAD and puADA in Oligonucleotides

No experimental features about the behavior of pyDAD and puADA (xanthosine) in oligonucleotides are known. Although computational simulations have been performed with the base pair in DNA (Florian & Leszczynski, 1995; Leach & Kollman, 1992), their results

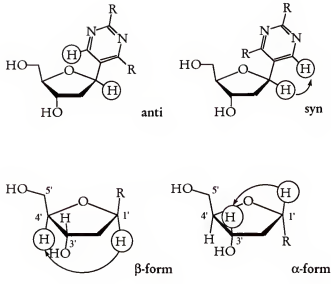


Figure 2-2. Structural analysis by NOE spectroscopy. Irradiation of the H-1' proton in a NOE experiment results is a characteristic pattern of responses from neighboring protons. In the β-form, irradiation of H-1' results in a signal enhancement at H-4'. In case of the α-form, irradiation of H-1' affects the signal intensity of H-3'. At the same time, the conformation of the nucleobase can be determined by the NOE effect between H-1' and H-6. In all cases, NOE effects are identified by difference NOE spectroscopy.

have little meaning in the absence of any experimental data. The only physicochemical studies performed with pyDAD and with a pyrazolopyrimidine analog of xanthosine, that also presents an acceptor-donor-acceptor hydrogen bonding pattern, were melting curves of 9-mers, containing a single internal nonstandard base pair (Piccirilli et al., 1990). These studies concluded that pyDAD and puADA form a proper base pair, though slightly less stable than duplex containing only standard bases. To collect further experimental data on nonstandard base pairs in oligonucleotides, two projects were started. First, in collaboration with Mary Kopka in the Dickerson laboratory at the University of California Los Angeles, work on the crystal structure of a modified Dickerson dodecamer, carrying 6-(puADA), 7-(pyDAD) substitutions, was initiated. Second, a new set of oligonucleotides with pyDAD/puADA substitutions in various positions was synthesized and their melting behavior was studied.

Melting studies of oligonucleotides

Perhaps the simplest way of determining the stabilizing or destabilizing effect of unusual nucleotides such as our nonstandard base pair on double-stranded DNA is to measure the DNA's melting temperature. A melting curve results from an increase in absorptivity (hyperchromicity) during the transition from double-stranded to single-stranded DNA. The denaturation of the DNA is therefore most easily followed by monitoring the UV absorbance of a sample while gradually increasing its temperature. Under ideal conditions, the system shows strongly cooperative melting behavior, which is indicated by a sharp, sigmoid transition. The melting temperature is defined as the temperature where the curve has a point of inflection, and can easily be determined from the first derivative of the curve. Besides being affected by environmental conditions such as the sample buffer, the melting temperature is sequence-dependent. Base stacking and hydrogen bonding between the two duplex strands, as well as the surrounding water can greatly influence the stability of double-stranded DNA. By studying standard oligonucleotides in parallel with homologous sequences that carry nonstandard base pairs, the contribution of the new building blocks in duplex stabilization can be measured.

The well-characterized dodecamer d(CGCGAATTCGCG) was chosen as the template sequence for our melting studies. Commonly known as the Dickerson dodecamer, its X-ray crystallographic structure [Drew, 1991 #1550], as well as a comprehensive set of melting studies by UV (Marky et al., 1983), NMR, and calorimetry (Patel et al., 1982) have been published.

Materials and Methods

Experimental Procedure

Reagents and solvents were purchased from Aldrich Chemicals and were of the highest available quality. Solvents were dried by distillation in the presents of the appropriate drying agents (Perrin & Armarego, 1988) and stored over activated molecular sieves (3-4 Å).

Instrumentation. Proton NMR spectra were recorded on an Varian Gemini at 300 MHz and 500 MHz, using TMS as internal reference. Carbon-13 NMR spectra were measured on a Varian Gemini at 75 MHz. Phosphorous-31 NMR were recorded on a Varian Gemini at 121 MHz, using phosphoric acid as an external standard. All mass spectroscopic analysis were performed by the departmental MassSpec Service (Dr. D.Powell) on a Finnigan MAT95Q (San Jose, CA). UV spectra were measured on a Varian Cary 1 Bio UV/VIS spectra-photometer. Column chromatography was performed with silica gel (200-425 MESH) from Fisher Scientific. For TLC, silica plates were purchased from Whatman (type AL SIL G/UV). The products were visualized by UV light at 254 nm and/or by submerging the plate in a Ce/Mo reagent (2.5% phosphormolybdic acid, 1% $Ce(IV)(SO_4)_2$ $4H_2O$, 6% sulfuric acid in water) followed by heating to visualize oxidazable compounds.

2,4-Diamino-5-(2',3'-isopropylidene-5'-O-trityl-β-D-ribofuranosyl)-pyrimidine (5). The starting materials **5a** and **5b** were synthesized by Andre Mueller at the ETH in Zurich (Switzerland), following the procedure by Piccirilli (Piccirilli, 1989). Either anomeric form (α (**5a**) and β (**5b**)) was extensively characterized for reference purposes. (**5b**) ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 7.84 (s, 1H, H-6), 7.39 (d, 6H, trityl ortho-H), 7.31 (t, 6H, trityl meta-H), 7.26 (q, 3H, trityl para-H), 5.37 (br, s, 1H, NH), 4.92 (q, 1H, H-3', $J_{H3/H4}$ = 6.0 Hz, $J_{H3/H2}$ = 7.0 Hz), 4.80 (br, s, 2H, NH), 4.78 (t, 1H, H-2', $J_{H2/H3}$ = 7.0 Hz, $J_{H2/H1}$ = 6.0 Hz), 4.60 (d, 1H, H-1'

$J_{H1/H2}$ = 6.0 Hz), 4.14 (q, 1H, H-4', $J_{H4/H3}$ = 6.0 Hz, $J_{H4/H5}$ = 2.5 Hz), 3.54 (dd, 1H, H-5'$_B$, $J_{H5/H4}$ = 2.5 Hz, $J_{H5/H5}$ = 10.5 Hz), 3.44 (dd, 1H, H-5'$_A$, $J_{H5/H4}$ = 2.5 Hz, $J_{H5/H5}$ = 10.5 Hz), 1.59 (s, 1H, Me), 1.37 (s, 1H, Me), ppm. ^{13}C-NMR (300 MHz, CDCl$_3$) δ: 177.0 (C-2), 162.6 (C-4), 162.0 (C-6), 155.0 (C-5), 143.1, 128.7, 128.0, 127.4 (trityl arom.), 115.3 (i-Pr), 105.2 (C(Ph)3), 87.3, 84.0, 83.7, 83.2, 81.7, 62.8 (sugar), 27.6 (i-Pr), 25.6 (i-Pr), ppm. ^{1}H-COSY (300 MHz, CDCl$_3$) δ: (H-3') 4.92/4.78, 4.92/4.14, (H-2') 4.78/4.92, 4.78/4.60, (H-1') 4.60/4.78/, (H-4') 4.14/4.92, 4.14/3.54, 4.14/3.44, (H-5'$_B$) 3.54/4.14, 3.54/3.44, (H-5'$_A$) 3.44/4.14, 3.44/3.54 ppm. NOE (300 MHz, CDCl$_3$) δ: IRR at 7.84 (H-6): 4.60 (H-1'), IRR at 4.60 (H-1') : 7.84 (H-6), 4.14 (H-4'), ppm. LRMS-FAB (m-NPA): (m/z) 525 (M^+ +1), 281 (M^+ -243), 243 (trityl). TLC: dichloromethane/diethyl ether 8:2 (10% methanol), R_f = 0.50.

(5a) ^{1}H-NMR (300 MHz, CDCl$_3$) δ: 7.78 (s, 1H, H-6), 7.42 (d, 6H, trityl ortho-H), 7.32 (t, 6H, trityl meta-H), 7.27 (d, 3H, trityl para-H), 5.46 (br s, 2H, NH), 5.03 (d, 1H, H-1', $J_{H1/H2}$ = 3.6 Hz), 4.81 (q, 1H, H-2', $J_{H2/H3}$ = 6.0 Hz, $J_{H2/H1}$ = 3.6 Hz), 4.75 (br s, 1H, NH), 4.70 (d, 1H, H-3', $J_{H3/H2}$ = 6.0 Hz), 4.35 (t, 1H, H-4', $J_{H4/H3}$ = 4.0 Hz), 3.34 (dd, 1H, H-5'$_B$, $J_{H5/H4}$ = 4.0 Hz, $J_{H5/H5}$ = 10.2 Hz), 3.26 (dd, 1H, H-5'$_A$, $J_{H5/H4}$ = 4.5 Hz, $J_{H5/H5}$ = 10.2 Hz), 1.50 (s, 1H, Me), 1.30 (s, 1H, Me), ppm. ^{13}C-NMR (300 MHz, CDCl$_3$) δ: 163.3 (C-2), 162.4 (C-4), 156.0 (C-6), 143.4 (C-5), 128.6, 128.0, 127.3 (trityl arom.), 112.5 (i-Pr), 102.6 (C(Ph)3), 87.4, 83.2, 82.9, 82.7, 82.1, 64.1 (sugar), 26.0 (i-Pr), 24.0 (i-Pr), ppm. ^{1}H-COSY (300 MHz, CDCl$_3$) δ: (H-1') 5.03/4.81, (H-2') 4.81/5.03, 4.81/4.70, (H-3') 4.70/4.81, 4.70/4.35, (H-4') 4.35/4.70, 4.35/3.34, 4.35/3.26, (H-5'$_B$) 3.34/4.35, 3.34/3.26, (H-5'$_A$) 3.26/4.35, 3.26/3.34, ppm. NOE (300 MHz, CDCl$_3$) δ: IRR at 5.03 (H-1'): 7,78 (H-6)/ 7.42 (trityl ortho-H)/ 4.81 (H-2'), IRR at 4.35 (H-4') : 4.70 (H-3'), ppm. LRMS-FAB (m-NPA): (m/z) 525 (M^+ +1), 281 (M^+ - 243), 243 (trityl). TLC: dichloromethane/diethyl ether 8:2 (10% methanol), R_f = 0.43

***N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(2',3'-isopropylidene-5'-O-trityl-β-D-ribofuranosyl)-pyrimidine (6).** Compound **5b** (2.1 g, 4 mmol) was evaporated twice with pyridine and dried under vacuum. The foamy, yellow residue was resolved in pyridine (40 mL) and catalytic amounts of DMAP were added. The solution was cooled to 0°C and *t*-butylbenzoyl chloride (2.4 mL, 12 mmol, 1.5 eq) was added dropwise to the clear, yellow solution. After a few minutes, a white compound started to precipitate and the suspension turned dark yellow. After 3 h at 0°C, the reaction was quenched by addition of methanol (1 mL). The ice bath was removed and the yellow suspension was shaken 10-15 min at ambient temperature. After evaporation in vacuo, the raw product was dried under vacuum. The raw product was purified over silica gel (ethyl acetate/hexane 1:2 and 1:1), yielding **6** (2.27 g, 67%). ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 9.54 (s, 1H, NH), 8.89 (s, 1H, NH), 8.79 (s, 1H, H-6), 7.79 (d, 4H, Bz ortho-H), 7.5-7.2 (m, 19H, Bz/trityl H), 4.90 (d, 1H, H-1', $J_{H1/H2}$ = 5.7 Hz), 4.7-4.6 (m, 2H, H-2', H-3'), 4.35 (m, 1H, H-4'), 3.45 (dd, 1H, H-5'$_B$, $J_{H5/H4}$ = 3.0 Hz, $J_{H5/H5}$ = 10.5 Hz), 3.33 (dd, 1H, H-5'$_A$, $J_{H5/H4}$ = 4.8 Hz, $J_{H5/H5}$ = 10.3 Hz), 1.38 (s, 3H, Me), 1.33 (s, 18H, t-Bu), 1.31 (s, 3H, Me), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 164.8, 164.7, 157.0, 156.7, 156.1, 155.7, 143.4, 131.3, 131.1, 128.5, 127.8, 127.7, 127.4, 127.1, 125.5, 125.4, 116.0, 115.7, 86.8, 84.3, 83.5, 82.3, 80.8, 63.6, 34.9, 31.0, 27.1, 25.4, ppm. ^{1}H-COSY (300 MHz, $CDCl_3$) δ: (H-1') 4.90/4.65, (H-2'/H-3') 4.65/4.90, 4.65/4.35, (H-4') 4.35/4.65, 4.35/3.40, (H-5'$_B$/H-5'$_A$) 3.40/4.35, ppm. NOE (300 MHz, $CDCl_3$) δ: IRR at 4.90 (H-1'): 8.79 (H-6) / 4.35 (H-4'), ppm. LRMS-FAB (m-NPA): (m/z) 845(M^+ +1). TLC: ethyl acetate/hexane 1:2, R_f = 0.15.

***N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(β-D-ribofuranosyl)-pyrimidine (7).** To a solution of **6** (1.35 g, 1.58 mmol) in methanol (10 mL) was added methanolic hydrogen chloride solution (2.5 mL, 10% w/w). After 90 min at RT, the TLC showed no more starting material. The reaction was stopped by bubbling argon gas through the solution (or degassing

on the evaporator) for 30 min. The clear orange solution was completely evaporated in vacuo and dried under vacuum overnight. The white-brown raw material was purified by column-chromatography over silica gel ($CHCl_3$/MeOH 11:1), giving **7** (640 mg, 64 %) as a white compound. ^{1}H-NMR (300 MHz, MeOH-*d*) δ: 8.67 (s, 1H, H-6), 7.92 (dd, 4H, ortho-H), 7.52 (q, 4H, meta-H), 4.91 (m, 1H, H-1'), 4.12 (m, 3H, H-2', H-3', H-4'), 3.79 (dd, 1H, H-5'$_B$, $J_{H5/H4}$ = 2.7 Hz, $J_{H5/H5}$ = 12.2 Hz), 3.70 (dd, 1H, H-5'$_A$, $J_{H5/H4}$ = 4.2 Hz, $J_{H5/H5}$ = 12.2 Hz), 1.39 (s, 9H, t-Bu), 1.32 (s, 9H, t-Bu), ppm. ^{13}C-NMR (300 MHz, MeOH-*d*) δ: 167.1, 167.0, 158.0, 157.9, 157.8, 157.4, 157.1, 132.2, 132.2, 129.0, 128.9, 126.8, 126.7, 118.1, 88.4, 79.3, 77.6, 73.2, 63.3, 35.9, 31.5, ppm. TLC: $CHCl_3$/MeOH 11:2, R_f = 0.38.

***N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(3',5'-O-[tetraisopropyldisiloxanyl]-β-D-ribofuranosyl)-pyrimidine (8).** Compound **7** (932 mg, 1.66 mmol) was evaporated three times with pyridine and dried under high vacuum overnight. The residue was re-dissolved in pyridine (8.8 mL) and tetraisopropyl-disiloxanedichloride (0.58 mL, 1.82 mmol, 1.1 eq.) was added to the clear, yellow solution. The mixture was shaken at RT. During the reaction, a white compound started to precipitate. The progress of the reaction was monitored by TLC. After 4 h, no starting material was detectable. The reaction solution was evaporated and dried under vacuum overnight. The residue was partitioned between chloroform and water (80 mL each) and extracted twice with chloroform and brine. The combined organic layers were dried over sodium sulfate and the solvent removed by evaporation. The product was purified by chromatography over silica gel (ethyl acetate/hexane 1:2) to yield **8** (917 mg, 60%) as a white, foamy compound. ^{1}H-NMR (500 MHz, $CDCl_3$) δ: 9.95 (br s, 1H, NH), 9.00 (br s, 1H, NH), 8.71 (s, 1H, H-6), 7.88 (dd, 4H, ortho-H), 7.47 (d, 4H, meta-H), 4.84 (d, 1H, H-1', $J_{H1/H2}$ = 5.5 Hz), 4.35 (t, 1H, H-3', $J_{H3/H2}$ = 6.5 Hz, $J_{H3/H4}$ = 6.5 Hz), 4.16 (m, 1H, H-2'), 4.14 (dd, 1H, H-5'$_B$, $J_{H5/H5}$ = 11.6 Hz, $J_{H5/H4}$ = 3.2 Hz), 3.99 (hex, 1H, H-4', $J_{H4/H5}$ = 3.1 Hz, 7.0 Hz, $J_{H4/H3}$ = 6.5

Hz), 3.93 (q, 1H, H-5'$_A$, $J_{H5/H5}$ = 11.6 Hz, $J_{H5/H4}$ = 7.0 Hz), 1.34 (s, 18H, t-Bu), 1.11-0.96 (m, 28H, i-Pr), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 164.9 (C=O), 156.8, 156.7, 156.3, 155.8 (pyrimidine), 131.4, 131.1, 127.6, 127.1, 125.6, 115.4, 84.3, 81.5, 75.3, 71.9, 62.6 (sugar), 35.0 (i-Pr), 31.1 (t-Bu), 17.5-17.0 (i-Pr), 13.5-12.5 (i-Pr), ppm. ^{1}H-COSY (300 MHz, $CDCl_3$) δ: (H-1') 4.84/4.16, (H-3') 4.35/4.16/3.99, (H-2') 4.16/4.35/4.84,(H-5'$_B$) 4.14/3.99/3.93, (H-4') 3.99/4.14/3.93, (H-5'$_A$) 3.93/4.14/3.99, ppm. TLC: hexane/ethyl acetate 2:1, R_f = 0.25.

***N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(3',5'-O-[tetraisopropyldisiloxanyl]-2'-O[{imidazole-1-yl}thiocarbanyl]-β-D-ribofuranosyl)-pyrimidine (9).** Starting material **8** (600 mg, 0.75 mmol) was dried under high vacuum overnight. The compound was dissolved in DMF (7 mL), and 1,1'-thiocarbonyl-diimidazole (306 mg, 1.72 mmol, 2 eq.) in DMF (3 mL) was added at RT. After 4 h, the reaction was finished according to TLC. The reaction was quenched by addition of ethyl acetate (50 mL) and water (12 mL). Following extraction, the organic layer was dried over sodium sulfate and the solvent evaporated. Overnight drying under high vacuum yielded **9** (740 mg, 108 %). The crude material was used directly for the next step. ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 8.56 (br s, 1H, NH), 8.32 (s, 1H, NH), 8.12 (s, 1H, H-6), 7.9 (m, 4H, Bz), 7.69 (s, 1H, imidazole), 7.60 (s, 1H, imidazole), 7.48 (m, 4H, Bz), 7.03 (s, 1H, imidazole), 6.14 (d, 1H, H-2', $J_{H2/H3}$ = 5.0 Hz), 5.55 (s, 1H, H-1'), 4.45 (q, 1H, H-3', $J_{H3/H2}$ = 5.0 Hz, $J_{H3/H4}$ = 8.7 Hz), 4.18 (m, 1H, H-5'$_B$, $J_{H5/H5}$ = 12.3 Hz), 4.02 (m, 2H, H-4', H-5'$_A$), 1.31 (s, 9H, t-Bu), 1.25 (s, 9H, t-Bu), 1.06-0.8 (m, 28H, i-Pr), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 82.8, 157.0, 156.0, 155.7, 155.6, 138.3, 136.9, 131.9, 130.7, 128.6, 127.7, 125.8, 125.0, 120.6, 118.0, 84.8, 81.7, 78.5, 70.3, 60.1, 35.1, 31.0, 17.4, 17.2, 17.1, 17.0, 16.7, 16.6, 13.2, 13.0, 12.6, 12.5, ppm. ^{1}H-COSY (300 MHz, $CDCl_3$) δ (H-2'): 6.14/5.55 (H-1') 6.14/4.45 (H-3'), (H-1'): 5.55/6.14 (H-2'), (H-3'): 4.45/6.14 (H-2') 4.45/4.0-4.3, (H-4', H-5'$_A$, H-5'$_B$), ppm. TLC: hexane/ethyl acetate 2:1, R_f = 0.20.

*N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(3',5'-O-[tetraisopropyldisiloxanyl]-2'-deoxy-β-D-ribofuranosyl)-pyrimidine (10). Compound 9 (740 mg, 0.81 mmol), dried under high vacuum overnight, was dissolved in toluene (10 mL). Separately, AIBN (97 mg, 0.59 mmol, 0.72 eq.) and tributyltin hydride (0.87 mL, 3.24 mmol, 4 eq.) were dissolved in toluene (5 mL). The two solutions were mixed and degassed with argon for 15-20 min at RT. The solution was then heated to 80°C. Gas generation was observed. TLC of the reaction mixture showed no more starting material after 60 min. The solution was cooled to RT and evaporated. The product was isolated by chromatography on silica gel (hexane/ethyl acetate 4:1), giving 10 (520 mg, 81%) as a white solid compound. ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 10.0 (br s, 1H, NH), 8.90 (br s, 1H, NH), 8.44 (s, 1H, H-6), 7.84 (dd, 4H, Bz), 7.47 (t, 4H, Bz), 5.17 (q, 1H, H-1', $J_{H1/H2A}$ = 8.4 Hz, $J_{H1/H2B}$ = 6.3 Hz), 4.50 (m ,1H, H-3'), 4.13 (dd, 1H, H-5'$_B$, $J_{H5/H4}$ = 3.6 Hz, $J_{H5/H5}$ = 11.5 Hz), 3.97 (m, 1H, H-4'), 3.75 (m, 1H, H-5'$_A$, $J_{H5/H4}$ = 8.7 Hz, $J_{H5/H5}$ = 11.5 Hz), 2.39 (m, 1H, H-2'$_B$), 2.30 (m, 1H, H-2'$_A$), 1.33 (s, 9H, t-Bu), 1.32 (s, 9H, t-Bu), 1.10 - 0.97 (m, 28H, i-Pr), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 164.8, 164.5, 157.3, 157.0, 156.2, 155.9, 131.1, 131.1, 127.5, 125.8, 125.6, 116.2, 87.3, 76.1, 73.3, 63.6, 53.5, 40.4, 35.1, 31.1, 25.2, 17.6, 17.4, 17.2, 17.0, 16.9, 13.4, 13.3, 12.9, 12.5, ppm. ^{1}H-COSY (300 MHz, $CDCl_3$) δ (H-1'): 5.17/2.35 (H-2'$_{A/B}$), (H-3'): 4.50/2.35 (H-2'$_{A/B}$) 4.50/3.97 (H-4'), (H-5'$_B$): 4.13/3.97 (H-4') 4.13/3.75 (H-5'$_A$), (H-4'): 3.97/4.50 (H-3') 3.97/4.13 (H-5'$_B$) 3.97/3.75 (H-5'$_A$), (H-5'$_A$): 3.75/3.97 (H-4') 3.75/4.13, (H-5'$_B$), ppm. LRMS-FAB (m-NPA): (m/z) 811 (M^+ +Na), 789 (M^+ +1). TLC: hexane / ethyl acetate 4:1, R_f = 0.10.

*N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(2'-deoxy-β-D-ribofuranosyl)-pyrimidine (11). Compound 10 (690 mg, 0.87 mmol) were dissolved in THF (4 mL). After addition of TBAF (1.74 mL, 1.74 mmol, 1 M in THF), the clear and slightly yellow solution was shaken at RT. After 45 min, TLC showed no more starting material. The reaction mixture was

evaporated and the residue purified by chromatography on silica gel ($CHCl_3$/MeOH 11:1), yielding **11** (423 mg, 89 %). ^{1}H-NMR (300 MHz, MeOH-*d*) δ: 8.48 (s, 1H, H-6), 7.95 (dd, 4H, Bz), 7.57 (dd, 4H, Bz), 5.29 (q, 1H, H-1', $J_{H1/H2B}$ = 5.1 Hz, $J_{H1/H2A}$ = 10.5 Hz), 4.32 (m, 1H, H-3'), 4.06 (m, 1H, H-4'), 3.73 (dd, 1H, H-5'$_B$, $J_{H5/H4}$ = 3.9 Hz, $J_{H5/H5}$ = 12.0 Hz), 3.64 (dd, 1H, H-5'$_A$, $J_{H5/H4}$ = 5.4 Hz, $J_{H5/H5}$ = 12.0 Hz), 2.38 (dd, 1H, H-2'$_B$, $J_{H2/H1}$ = 5.1 Hz, $J_{H2/H2}$ = 13.0 Hz), 2.20 (hex, 1H, H-2'$_A$, $J_{H2/H3}$ = 6.0 Hz, $J_{H2/H2}$ = 13.0 Hz, $J_{H2/H1}$ = 10.5 Hz), 1.36 (s, 18H, t-Bu), ppm. ^{13}C-NMR (300 MHz, MeOH-*d*) δ: 167.2, 167.1, 158.5, 158.2, 158.0, 157.7, 157.5, 132.4, 132.0, 128.9, 126.9, 126.7, 119.4, 89.9, 77.1, 73.5, 63.6, 54.2, 41.8, 36.0, 35.9, 31.5, ppm. ^{1}H-COSY (300 MHz, MeOH-*d*) δ: (H-1'): 5.29/2.30 (H-2'$_{A/B}$), (H-3'): 4.32/2.30 (H-2'$_{A/B}$) 4.32/4.06 (H-4'), (H-4'): 4.06/4.32 (H-3') 4.06/3.65 (H-5'$_{A/B}$), (H-5'$_{A/B}$): 3.65/4.06 (H-4'), (H-2'$_{A/B}$): 2.30/5.29 (H-1') 2.30/4.32 (H-3'), ppm. NOE (300 MHz, MeOH-*d*) δ: IRR at 5.29 (H-1'): 8.48 (H-6) / 4.06 (H-4') / 2.30 (H-2'), IRR at 4.06 (H-4') : 5.29 (H-1')/4.32 (H-3')/3.65 (H-5'), ppm. LRMS-FAB (m-NPA): 569 (M^+ +Na), 547 (M^+ +1). TLC: $CHCl_3$/MeOH 11:1, R_f = 0.22.

[2'-3H]-N2,N4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(2'-deoxy-β-D-ribofuranosyl)-pyrimidine (11a). Compound **9** (100 mg, 0.11 mmol) was dissolved in AIBN (13.5 mg, 0.082 mmol, 0.75 eq) and toluene (3.5 mL). After degassing for 15 min with argon, [^{3}H] tri-*n*-butyltin hydride (60 μL, 0.22 mmol, 2 eq) was added and the reaction mixture heated to 80°C. After 2h, TLC indicated about 50% starting material. More tri-*n*-butyltin hydride (60 μL) was added, and the reaction was heated for another 2 h at 80°C. The reaction was stopped by evaporation of the solvent and the oily residue was dried under vacuum. The raw material was dissolved in THF (1 mL, anhydrous) and TBAF (0.25 mL, 1 M in THF) was added. After 1 h at RT, the solvent was evaporated and the crude material purified over silica gel ($CHCl_3$/MeOH 11:1), yielding **11a** (43 mg, 75% based on **9**).

*N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(5'-O-(4',4''-dimethoxytrityl)-2'-deoxy-β-D-ribofuranosyl)-pyrimidine (12). Compound **11** (423 mg, 0.77 mmol) was coevaporated with pyridine twice. The foamy residue was then dissolved in pyridine (5 mL). The slightly yellow solution was treated with dimethoxytrityl chloride (313 mg, 0.92 mmol) dissolved in pyridine (5 mL). The solution was shaken at RT for 4-5 h. The reaction was stopped by addition of water (0.3 mL). After evaporation, the residue was partionated between ethyl acetate (40 mL) and saturated $NaHCO_3$ solution (40 mL). The combined organic layers were dried over sodium sulfate and filtered. After evaporation of the solvent, the residue was purified by chromatography on silica gel ($CHCl_3$/MeOH 20:1), giving **12** (495 mg, 76 %) as a white, foamy product. ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 10.6 (br s, 1H, NH), 9.90 (br s, 1H, NH), 8.45 (s, 1H, H-6), 7.95 (d, 2H, Bz), 7.76 (d, 2H, Bz), 7.42-7.05 (m, 13H, arom.), 6.65 (d, 4H, arom.), 5.36 (q, 1H, H-1', $J_{H1/H2A}$ = 5.2 Hz, $J_{H1/H2B}$ = 9.3 Hz), 4.74 (m, 1H, H-3'), 4.49 (m, 1H, H-4'), 3.64 (s, 6H, -OCH3), 3.28 (m, 2H, H-5'$_{B/A}$), 2.66 (m, 1H, H-2'$_B$), 2.45 (m, 1H, H-2'$_A$), 1.31 (s, 9H, t-Bu), 1.24 (s, 9H, t-Bu), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 164.9, 164.5, 158.3, 157.5, 157.0, 156.5, 155.8, 149.5, 144.4, 135.9, 135.5, 131.5, 130.4, 129.8, 127.9, 127.6, 127.3, 126.6, 125.8, 125.3, 123.6, 116.7, 112.9, 87.5, 86.1, 77.2, 75.7, 73.0, 64.3, 54.9, 39.9, 34.9, 34.8, 31.0, ppm. TLC: CHCl3/MeOH 11:1, R_f = 0.61.

*N*2,*N*4-Di-(*t*-butyl-benzoyl)-2,4-diamino-5-(5'-O-(4',4''-dimethoxytrityl)-2'-deoxy-β-D-ribofuranosyl)-pyrimidine-3'-O-cyanoethoxydiisopropyl-phophoramidite (13). Compound **12** (101 mg, 0.12 mmol) was dissolved with DMAP (10 mg) in acetonitrile (1.3 mL). Diisopropyethylamine (84 μL, 0.48 mmol) was added and the suspension cooled in an ice bath. Finally, cyanoethyl-diisopropyl-chlorophosphoramidite was added dropwise and the now clear solution was shaken at temperature for 15 min. The reaction was stopped by partitioning the mixture between saturated $NaHCO_3$ solution (25 mL) and dichloromethane

(25 mL). The combined organic layers were dried over sodium sulfate, filtered, and the filtrate evaporated. The raw material was purified on silica gel (CH_2Cl_2/EtOAc/TEA 70:30:1), yielding **13** (110 mg, 90%). ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 10.1 (br s, 1H, NH), 8.76 (br s, 1H, NH), 8.51 (d, 1H, H-6), 7.86 (dd, 2H, Bz), 7.50-7.10 (m, 13H, arom.), 6.75 (d, 4H, arom.), 5.27 (m, 1H, H-1', $J_{H1/H2A}$ = 5.4 Hz), 4.53 (m ,1H, H-3', $J_{H3/HX}$ = 5.4 Hz), 4.35 (m, 1H, H-4'), 3.73 (s, 6H, -OCH_3), 3.75-3.50 (m, 4H, -CH_2CH_2CN), 3.28 (m, 2H, H-5'B/A), 2.64 (m, 1H, H-2'B), 2.53 (t, 1H, i-Pr), 2.45 (t, 1H, i-Pr), 2.33 (m, 1H, H-2'A), 1.34 (s, 18H, t-Bu), 1.25-1.05 (m, 12H, i-Pr), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 177.0, 164.6, 164.5, 164.3, 158.4, 157.9, 157.8, 157.0, 156.3, 156.2, 155.9, 144.4, 135.6, 131.3, 130.8, 129.9, 128.0, 127.8, 127.5, 127.4, 126.8, 125.7, 125.6, 117.8, 117.2, 113.0, 86.3 (d), 75.8 (d), 74.9 (d), 63.8 (d), 58.3 (d), 58.1(d), 55.1, 43.1 (dd), 38.8, 35.0, 31.0, 24.5, 20.2 (dd), ppm. ^{31}P-NMR (300 MHz, $CDCl_3$) δ: 146.8, 145.8, ppm (against phosphoric acid in deuterated water as external standard). LCMS FAB (m-NPA): 1049 (M^+ +1). TLC: $CHCl_3$/MeOH 11:1, R_f = 0.61.

2,4-Diamino-5-(2'-deoxy-β-D-ribofuranosyl)-pyrimidine (15). Compound **11** (143 mg, 0.22 mmol) was dissolved in methanol (1 mL). Ammonia solution (12 mL, 30%) was added and the tightly closed sample was heated to 60°C overnight. After cooling to RT, the sample was lyophilized and the raw material purified on silica gel ($CHCl_3$/MeOH 3:2) yielding **15** (27 mg, 51%). To form the HCl-salt, the sample was extracted twice with HCl (0.01 M) against hexane. The aqueous layer was lyophilized, yielding **15** *(HCl-form)* (30 mg, 45%) as a white foamy solid. ^{1}H-NMR (300 MHz, MeOH-*d*) δ: 7.71 (s, 1H, H-6); 4.98 (q, 1H, H-1', $J_{H1/H2B}$ = 5.4 Hz, $J_{H1/H2A}$ = 11.1 Hz); 4.40 (m, 1H, H-3', $J_{H3/H2A}$ = 5.7 Hz); 3.95 (q, 1H, H-4', $J_{H4/H5A/B}$ = 2.6 Hz); 3.74 (oct, 2H, H-5', $J_{H5B/H5A}$ = 12.0 Hz, $J_{H5/H4}$ = 2.6 Hz); 2.21 (oct, 1H, H-2'$_A$, $J_{H2A/H2B}$ = 13.2 Hz, $J_{H2A/H1}$ = 11.5 Hz, $J_{H2A/H3}$ = 6.5 Hz); 2.00 (oct, 1H, H-2'$_B$, $J_{H2B/H2A}$ = 13.2 Hz, $J_{H2B/H1}$ = 5.5 Hz, $J_{H2B/H3}$ = 1.0 Hz), ppm. ^{13}C-NMR (300 MHz, MeOH-*d*) δ: 164.8, 140.3, 109.9, 89.6,

77.7, 74.3, 62.9, 41.0, ppm. LRMS FAB (m-NPA): 227 ($M^+ + 1$). TLC: chloroform/methanol 3:2, R_f = 0.32.

2'-Deoxyxanthosine-5'-triphosphate (16). 2'-Deoxyguanosine triphosphate (sodium-salt, 10 mg, 16.7 mmol) was dissolved in a solution of water (220 μL) and sodium sulfite (10 mg, 80 mmol). A mixture of HCl (8.7 μL, 2 M) and acetic acid (glacial, 25 μL) was added and the sample incubated at RT for 3 h. The reaction was quenched with Tris base (400 μL, 1 M). The raw material can be stored at -20°C before being purified by RP-HPLC (Nova Pak C-18 Radial Pak cartridge (Waters), 25 x 100 mm, TEAA (100 mM, pH 7), linear gradient to 10% acetonitrile over 25 minutes). The combined product fractions were lyophilized and the residue was dissolved in Tris-HCl (2 mL, 10 mM, pH 7.0). The yield of **16** was determined by UV absorbance (4.2 mg, 42%, λ_{max}: 247/277 nm, ε: 10,000/9100 $M^{-1}cm^{-1}$). The purity of the material was > 97% as determined by analytical RP-HPLC (same system as above) and anion-exchange HPLC (Macrosphere 300A WAX 7U (Alltech, Deerfield IL) 4.6 mm x 250 mm; solvent A = water; solvent B = TEA-bicarbonate (0.8 M, pH 7.2); curved gradient (#7) from 1% to 50% B in 15 min).

[8-^{3}H] 2'-Deoxyxanthosine-5'-triphosphate (16a). The procedure used for **16** is also applicable for radiolabelled material. [8-^{3}H]-dGTP (Na-salt, 200 μL, 16 Ci/mmol in water/EtOH 1:1) was lyophilized and the residue redissolved in water (5.5 μL). After addition of sodium sulfite solution (1 μL, 5 M), the reaction was initiated by adding HCl (0.2 μL, 2 M) and glacial acetic acid (3.24 μL). After incubation at 30°C for 30 min, the reaction was quenched with Tris base (36 μL, 1 M) and purified by RP-HPLC as described for **16**. A significant amount of starting material was collected as a separate fraction and was recycled after lyophilization. The yield of **16a** was determined by UV absorbance (1.82 nmol, 29.1 μCi, 13.3%).

Oligonucleotide synthesis with pyDAD/puADA. DNA synthesis was performed on an Applied Biosystems (ABI) DNA synthesizer model 394 at the 0.2 to 10-μmol scale. The columns were purchased from ABI (500 Å, CPG solid support). The standard reagents were purchased from ABI/Perkin Elmer. All reactants (phosphoramidites) were diluted to half the recommended concentration.

The coupling steps were run according to the standard protocol with double coupling time for the nonstandard base and the material was detritylated at the end of the synthesis. The coupling efficiency, calculated based on the trityl release during the solid-phase synthesis, indicated an efficiency of > 95% for **13**. The detritylated oligonucleotide was cleaved from the solid support and dried in the speed-vac. In oligonucleotides that contain no 2'-deoxyxanthosine, the residue was dissolved in concentrated ammonia and incubated at 60°C overnight.

Sequences that contain 2'-deoxyxanthosine were treated with DBU prior to ammonia deprotection. For a typical 0.2 μmol scale synthesis, the dry, base-protected oligonucleotide was dissolved in pyridine (1 mL, anhydrous) and mixed with DBU (100 μL). After overnight incubation at 60°C, the pyridine was evaporated and the oily residue redissolved in water (250 μL). The oligonucleotide was recovered by ethanol precipitation or dialysis (Spectrum/Por® Membrane MWCO: 1,000; 24 mm diameter) if the sequence was shorter than 20 base pairs, and was subsequently subjected to ammonia treatment.

Oligonucleotide purification. The deprotected oligonucleotide was purified by preparative denaturing PAGE (12-20%.) The product band was excised and the pieces extracted with "crush&soak" buffer (Appendix C10). The concentration of the recovered oligonucleotide was determined by UV absorbance. Further, a small sample was 5'-end labelled with [γ-^{32}P]-ATP (Appendix C7) and analyzed by PAGE.

Successful incorporation of pyDAD and puADA was shown through enzymatic digestion of the oligonucleotides with phosphodiesterase (*Crotalus durissus terrificus*) and alkaline phosphatase (*Bovine calf intestine*). For details see appendix C11. Subsequent RP-HPLC analysis (5% acetonitrile in TEAA (10 mM, pH 7) over 15 min, then 10% acetonitrile within 8 min) was used to quantify the composition of the oligonucleotides (pyDAD; $\lambda_{max.}$ = 281 nm, ε = ~6000 $M^{-1}cm^{-1}$, t_R = 23-24 min; puADA; $\lambda_{max.}$ = 250/278 nm, ε = 10000/9100 $M^{-1}cm^{-1}$, t_R = 7-8 min).

Synthesis and purification of Dickerson dodecamer for X-ray crystallography. The Dickerson dodecamer analog was prepared on a 10-μmol scale. Synthesis conditions were as previously described. The oligonucleotide was deprotected by DBU and ammonia treatment, followed by purification on preparative polyacrylamide gels (20%, 7 M urea). Product bands were isolated by crush&soak and the salt concentration reduced by dialysis (Spectrum/Por® Membrane, MWCO: 1,000; 24 mm). After lyophilization, the residue was dissolved in water and purified by gel filtration (Pharmacia NAP-25 columns). The product-containing fractions were pooled and lyophilized, yielding approximately 400 OD_{260} (~14 mg, 3.33 μmol, 37%). Analytical samples were analyzed by PAGE (20%, 7 M urea, 5'-^{32}P labelled samples), anion-exchange HPLC (Macrosphere 300A WAX 7U (Alltech, Deerfield IL) 4.6 mm x 250 mm; solvent A = Tris-HCl (25 mM, pH 8) + 10% acetonitrile; solvent B = Tris-HCl (25 mM, pH 8), NaCl (2 M) + 10% acetonitrile; linear gradient from 2% to 10% B in 20 min), and anion-exchange FPLC (Pharmacia MonoQ 5/5 (Pharmacia, Piscataway NJ); solvent A = NaOH (50 mM, pH 12.5); solvent B = NaOH (50 mM, pH 12.5), NaCl (1 M); linear gradient from 20% to 100% B in 30 min). The final purity of the oligonucleotide was greater than 95%. The material was submitted to our collaborator, Mary Kopka in the Dickerson laboratory at University of California, Los Angeles.

Melting curve studies. Oligonucleotides were dissolved in Breslauer buffer (10 mM sodium phosphate and 1 mM EDTA, adjusted to pH 7.0 and the desired overall NaCl concentration) (Vesnaver & Breslauer, 1991). Prior to use, the buffer was sterile filtered and degassed for at least 30 min. The strand concentrations were determined at 260 nm (25°C) and were in the range of 3 to 55 μM. Optical melting curves were measured on a Varian Cary 1 Bio UV/VIS Spectrophotometer, equipped with a 12 cell sample changer and Peltier temperature control element. Quartz cuvets (1-cm pathlength) were used for oligonucleotide concentrations <20 μM while experiments at higher concentrations were performed in special cells with 0.1-cm pathlength. Samples were cycled between 10 to 85°C with 0.5 - 1°C/min. Data were recorded and analyzed, using the Cary UV Win v.3.0 software. The melting temperature was defined at the peak in a first derivative plot of the melting curve.

Results and Discussion

The Chemistry of pyDAD and puADA

The two building blocks of the pyDAD/puADA base pair, 2,4-diamino-2'-deoxy-ribofuranosyl-pyrimidine and 2'-deoxyxanthosine, were prepared for automated DNA synthesis and enzymatic studies (Chapter 3 and 5).

2'-Deoxyxanthosine: the puADA nucleoside

The phosphoramidite of 2'-deoxyxanthosine was prepared from 2'-deoxyguanosine. A shorter synthesis starting directly with xanthosine was also considered. Lower overall yields

and unidentified side-products lead us to abandon this route, however. Both approaches were studied in depth by Simona Jurczyk (Sulfonics, Alachua FL).

The synthesis, starting from 2'-deoxyguanosine, yielded DNA-synthesis-grade material in up to 20% overall yield. The stereochemistry was confirmed by NOE experiments (S. Jurczyk, unpublished). The triphosphate of 2'-deoxyxanthosine and its [8-^{3}H]-labelled analog were prepared by a simple one-step deamination of 2'-deoxyguanosine triphosphate (Piccirilli, 1989).

2,4-Diamino-5-(2'-deoxyribofuranosyl)-pyrimidine: the pyDAD nucleoside

As part of the synthetic efforts towards the expanded genetic alphabet, improvements in the preparation of pyDAD were studied. The biggest losses in yield over the entire pathway were suffered in step 5, the protection of the exocyclic amino-groups on the pyrimidine ring, and in step 6, the deprotection of the isopropylidene and trityl group (Figure 2-1).

Analysis of the steps identified two reasons for the drop in yields. In step 5, Piccirilli's conditions (RT, 2 h, 10 eq. Bz chloride) (Piccirilli, 1989) generated six to seven product spots. Although partial hydrolysis with ammonia at the end of the synthesis could convert some of the tetrasubstituted side-product **21** back into the disubstituted product **6**, the desired product was isolated in a maximum of 50% yield. In step 6, the cleavage of the sugar protection groups results in the very hydrophilic compound **7** (R = Bz) which could hardly be purified over silica gel, reducing the yields to 40%. A summary of possible side-reactions is shown in Figure 2-3.

Optimization of step 5 was possible by modification of the reaction conditions. After a series of experiments, varying the reaction temperature and the stoichiometry of protecting reagent with or without catalysts (DMAP, TEA), the yield was raised to 70%. Reacting three

mol-equivalents of protecting reagent (*tert* butylbenzoyl chloride or benzoyl chloride) at 0°C in the presence of catalytic amounts of DMAP not only reduced the amount of side-product but also simplified the purification of intermediate **6** substantially. At temperatures below RT, increasing amounts of *tert* butylbenzoyl chloride or benzoyl chloride in the reaction mixture (up to 20 eq) led to the preferred formation of the tetra-substituted product **21** in up to 80% yields. The product has a significantly higher mobility in the TLC system (R_f = 0.7, compared

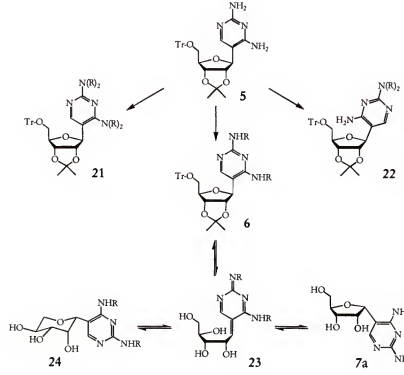


Figure 2-3. Possible side-reactions and isomerization during step 5 and 6. Tetra-substituted product **21**, formed in the presence of large excess reagent, as well as **2,2**-disubstituted material **22** are likely side-products to the **2,4**-disubstituted product **6** in step 5. Under the acidic conditions in step 5 and 6, the sugar of all three products (**6**, **21**, **22**) can be protonated, which could lead to isomerization product **23**, the α-anomer **7a**, as well as the pyranosyl derivative **24**.

to 0.15 for **6** (ethyl acetate/hexane 1:3)) and was clearly identified by integration of the proton signals in NMR experiments. The structural integrity of the ribose was confirmed by NOE experiments. Attempts to hydrolyze the isopropylidene and trityl-groups on **21** failed. TLC indicated a complex mixture of products. In addition, the partial hydrolysis with ammonia to convert the tetra-substituted material into the disubstituted product had only limited success. TLC indicated the formation of two new major products, the first being **6**, the second being a new side-product with a slightly higher Rf-value of 0.25 (ethyl acetate/hexane 1:2).

The new side-product was observed exclusively at elevated reaction temperatures. Exploring the effect of the reaction temperature from 0°C to 100°C, increasing amounts of the unknown compound were observed above RT. Proton-NMR of the purified material indicated the correct number of protons for the disubstituted product. Based on proton NMR, NOE, and UV data, the 2,2-disubstituted configuration (**22**, Figure 2-3) was proposed for the new product. At temperatures above 50°C, **22** was formed almost exclusively, even in the presence of 20 equivalents of benzoyl chloride reagent. The formation of tetra-substituted side-product **21** was not observed. We suggest that the 2,2-disubstitution of **5** is thermodynamically favorable over the formation of the 2,4-diprotected material. Furthermore, the bulky protection group probably blocks further substitution by rotating the pyrimidine base around the C-glycosidic bond, making the remaining 4-amino group inaccessible.

Another factor that causes problems during the purification of the reaction product in step 5 was the excess of reagent. Although already reduced under the new conditions, quenching the reaction with water turned the remaining reagent into its benzoic acid derivative and aqueous HCl. While the acidic conditions may cause partial isomerization of the products (Figure 2-3, **23**), the benzoic acid makes purification of the raw material by silica gel chromatography almost impossible. Beside reducing the load of benzoic acid by minimizing

the excess of reagent in the first place, the quenching of the reaction with methanol further simplifies the purification by reducing the risk of isomerization and by formation the methyl ester of benzoic acid which is separable by column chromatography.

For step 6, the problem of hydrophilicity of the products was overcome by using the more hydrophobic *tert* butylbenzoyl protection group on the nucleobase. Switching the reagent from benzoyl chloride to *tert* butylbenzoyl chloride did not affect the efficiency of the protection on the nucleobase. The deprotected ribose **8** could be purified without further problems. The average yield of this step was increased from 40 to 70%.

In summary, the reported modification of the reaction conditions (catalytic amounts DMAP, minimal excess reagent, low reaction temperature) and the usage of *tert* butylbenzoyl chloride to protect the exocyclic amino-groups in step 5 have helped to improve the overall yield of the pyDAD-phosphoramidite synthesis from 1.5 to 4%. We assume that these modifications to step 5 and 6 not only raise the yields of these particular steps but also improve the quality of intermediate 7. The better quality of the starting material for the deoxygenation could explain the observed twofold increase in the overall yield for these steps. Finally, deprotection studies on the pyDAD nucleoside **11** indicate complete removal of the *tert* butylbenzoyl group under standard oligonucleotide deprotection conditions (ammonia, 60°C, overnight).

Structural studies on 2,4-diamino-5-(2'-deoxyribofuranosyl)-pyrimidine

The structural integrity of pyDAD has repeatedly been a factor of uncertainty, especially in situations where unexpected results from melting studies or enzymatic experiments were obtained. The last structural study of a pyDAD derivative was the X-ray crystallographic structure of the ribonucleoside 7, determined by T.Krauch (T.Krauch,

S.A.Benner – unpublished results). The preservation of the correct stereochemistry during the following multi-step synthesis, leading to the 2'-deoxynucleoside needed to be addressed. Furthermore, the stability of the nonstandard nucleoside in an oligonucleotide during deprotection with ammonia was uncertain.

The stereochemistry of key intermediates from **5** to **15** was determined by difference NOE spectroscopy. First, the verification of this method was obtained by measuring the NOE effects in the two reference samples **5a** and **5b** (see Materials and Methods). The peak patterns for the α-form (**5a**), as well as for the β-form (**5b**) were precisely as predicted. Irradiation of H-1' in **5a** results in an increase of signal intensity for H-2' and H-3', as well as the more distant protons in the trityl group. For **5b**, the irradiation of H-1' causes a signal enhancement in H-4'. In both experiments, a NOE effect was observed for H-6, suggesting that the favored conformation of the nucleobase around the C-glycosidic bond is syn. These data agree with previous reports of preferred syn-conformation of C-nucleosides (O'Leary & Kishi, 1994) and the crystal structure data from Krauch.

In our experiments, each sample was analyzed by irradiation in at least two frequencies to confirm the data and structure of the nucleoside. As an example, the spectra for **11** are shown in Figure 2-4. NOE experiments for the intermediates **6**, **11**, and **15** all indicated the correct configuration at the anomeric center. These data are supported by previous work by Townsend and coworkers (Singh et al., 1991; Wise et al., 1978) on the stereochemistry of pseudouridine under deoxygenation conditions.

Another step where the nucleoside might epimerizes is the deprotection of the oligonucleotides after successful DNA synthesis. The stability of pyDAD at elevated temperature under strong basic conditions over an extended period of time needed to be confirmed. To simulate conditions normally encountered during deprotection of

oligonucleotides, **11** was incubated with DBU/pyridine at 60°C overnight. Following lyophilization, the sample was then incubated with concentrated ammonia solution at 60°C overnight. Neither treatment showed any indication of racemization at C-1', according to NOE measurements.

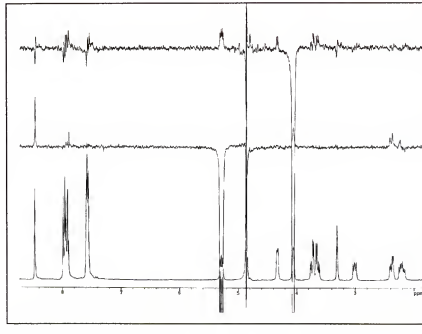


Figure 2-4. NOE experiment on **11** confirm the β-conformation of the nucleotide. The bottom spectrum shows the regular proton NMR. In the middle, irradiation at 5.15 ppm (H-1') results in a signal enhancement in H-6 (8.22 ppm), H-4' (4.02 ppm) and one of the H-2' protons (2.4 ppm). On the top, the experiment was repeated by irradiation at 4.02 ppm (H-4'). Beside the expected effect on the signal of H-1' (5.15 ppm), smaller peaks for H-3' (4.25 ppm) and H-5' (3.70 ppm) can be observed. The peaks to the left are artifacts, resulting from phase correction.

The pyDAD/puADA Nonstandard Base Pair in Oligonucleotides

Two separate projects were initiated to collect more information about the structural effects of nonstandard nucleotides in DNA. Milligram quantities of the [6-puADA, 7-pyDAD] Dickerson dodecamer (Table 2-1) were synthesized and purified for X-ray crystallography. In

addition, a series of oligonucleotides containing pyDAD and puADA, mismatches, or wobble base pairs in the same position were synthesized to study base pairing properties by UV melting curves (Table 2-1).

Crystal structure of pyDAD/puADA in an oligonucleotide. The most detailed information about the structural effects of the nonstandard base pair on nucleic acid could be acquired from a high-resolution X-ray crystallographic structure of an oligonucleotide containing pyDAD and puADA. In a large-scale DNA synthesis, milligram quantities of the [6-puADA, 7-pyDAD] Dickerson dodecamer (DiDo) were prepared and purified. Crystallization experiments, set up by our collaborator Mary Kopka in the Dickerson laboratory at the University of California in Los Angeles, have only yielded spherulites, which were not suitable for diffraction studies.

Table 2-1. Oligodeoxyribonucleotide dodecamers for UV melting studies.

Description	Abbreviation	Sequence
[6-puADA, 7-pyDAD] DiDo	[6X,7k] DiDo	5'-CGCGAXkTCGCG-3'
[6-pyDAD, 7-puADA] DiDo	[6k,7X] DiDo	5'-CGCGAkXTCGCG-3'
[6-pyDAD, 7-pyDAD] DiDo	[6k,7k] DiDo	5'-CGCGAkkTCGCG-3'
[6-puADA, 7-puADA] DiDo	[6X,7X] DiDo	5'-CGCGAXXTCGCG-3'
[5-pyDAD, 8-puADA] DiDo	[5k,8X] DiDo	5'-CGCGkATXCGCG-3'
Reference DiDo	DiDo	5'-CGCGAATTCGCG-3'
[6-T, 7-T] DiDo	[6T,7T] DiDo	5'-CGCGATTTCGCG-3'
[6-T, 7-C] DiDo	[6T,7C] DiDo	5'-CGCGATCTCGCG-3'
[5-T, 8-C] DiDo	[5T,8C] DiDo	5'-CGCGTATCCGCG-3'
[6-G, 7-T] DiDo	[6G,7T] DiDo	5'-CGCGAGTTCGCG-3'
[5-G, 8-T] DiDo	[5G,8T] DiDo	5'-CGCGGATTCGCG-3'

Dickerson dodecamer analogs, carrying pyDAD/puADA base pairs in different positions in the oligonucleotide. The position of pyDAD is marked by k, while X represents puADA. The melting temperatures of the pyDAD/puADA sequences were compared to the reference sequence, as well as to a set of T/T and C/T mismatches and G/T wobble base pairs.

Melting studies of oligonucleotides

The effect of pyDAD/puADA base pairs in various positions in a dodecamer on the melting temperatures (T_M) of the oligonucleotides were investigated. A series of sequences was synthesized and purified by PAGE (Table 2-1). Melting experiments were performed with the oligonucleotides at various salt concentrations, as well as under variable pH conditions.

Melting temperatures at high salt concentrations. The stabilization of duplex DNA by the pyDAD/puADA base pair was studied by comparing the melting temperatures to analogous sequences carrying either a standard AT base pair, a pyrimidine-pyrimidine mismatch, or a GT wobble base pair in the same position. The experiments were performed at high salt concentrations (1 M NaCl). Under these conditions, structural effects that arise from the self-complementary nature of the sequence are effectively suppressed (Marky et al., 1983). The apparent monophasic transition indicated cooperative duplex-to-single-strand conversion of the dodecamer. As listed in Table 2-2, the presence of two nonstandard base pairs in the dodecamer caused a sequence-dependent drop of 10 to 15°C in the melting temperature of the modified oligonucleotides.

In comparison, neither of the Dickerson dodecamer-analogs with the TT and TC mismatches (Table 2-1) showed cooperative melting. In a pyrimidine-pyrimidine mismatch, the two bases are approximately 2Å further apart from each other than a regular pyrimidine-purine base pair, making the formation of hydrogen bonds improbable (Saenger, 1984).

Finally, the introduction of a G/T wobble base pair in the central region of the Dickerson dodecamer sequence reduced the melting temperature by 30 to 35°C. "Wobbling" is the result of steric misalignment of bases in a base pair The GT base pair can form in the presence of rare tautomeric forms of either thymidine or guanosine and is considered the overall "most stabilizing" misaligned base pair combination (Saenger, 1984).

Table 2-2. Melting temperatures for Dickerson dodecamers, containing nonstandard base pairs

Oligonucleotide	T_M (°C)
Reference DiDo	66.5 (0.7)
[6X,7k] DiDo	55.5 (1.0)
[6k,7X] DiDo	54.9 (1.4)
[6k,7k]/[6X,7X] DiDo	49.3 (1.7)
[5k,8X] DiDo	50.9 (0.7)
[6T,7T] DiDo	not detected
[6T,7C] DiDo	not detected
[5T,8C] DiDo	not detected
[6G,7T] DiDo	~30 (weak)
[5G,8T] DiDo	35.5 (1.0)

All sequences were dissolved in Breslauer's buffer (10 mM sodium phosphate (pH 7), 1 mM EDTA) at 1 M NaCl. The total oligonucleotide concentration was adjusted to 5 μM ($\pm$0.5 μM). Samples were cycled between 10-85°C for at least 5 rounds. The first cycle was not considered for the calculated T_M's. The reported melting temperatures are the averaged data from the remaining cycles. The standard deviations are given in parentheses.

From all the experimental data, we conclude that, despite the lower melting temperature for the pyDAD/puADA dodecamer in comparison to the reference sequence, the nonstandard base pair contributes to the overall stability of the oligonucleotide by formation of hydrogen bonds. The precise causes for this observed reduction of the melting temperature for oligonucleotides, containing pyDAD/puADA base pairs, can only be speculated. Steric factors as well as electrostatic interactions may affect the duplex formation. However, melting curves are not sophisticated enough to address these questions. Furthermore, the sequence-dependence of the melting temperature should not be forgotten. Kawase and coworkers (Kawase et al., 1986) studied the effect of sequence context on the T_M of a dodecamer. Simply switching AT and GC base pairs changed the T_M by 17°C (Table 2-3).

Table 2-3. Melting temperatures of self-complementary dodecamers, report by (Kawase et al., 1986)

d(GGGAAXYTTCCC) XY	T_M (0.5 A_{260})
GC	59.2°C
CG	51°C
AT	54.8°C
TA	42.3°C

The melting temperatures were determined in 10 mM cacodylate, 0.1 M NaCl. The variations of the melting temperatures for this sequence suggest, that the 10 to 15°C drop upon incorporation of pyDAD and puADA may only be the result of sequence-dependent fluctuations.

pH-dependency of melting temperatures. The effect of pH changes on the melting temperature of oligonucleotides containing nonstandard base pairs was investigated. At neutral pH, approximately 50% of pyDAD is protonated (pK_A (N-1/3)= 6.7) and >90% of puADA is deprotonated (pK_A = 5.7). Placing these charged residues next to each other in an oligonucleotide may account for the observed lower melting temperature. As for the previous salt-dependency studies, the reference dodecamer and the [6X,7k] DiDo were studied and the results listed in Table 2-4.

While the melting temperature of the reference oligonucleotides seemed largely unaffected by the pH changes, the presence of pyDAD and puADA further destabilized the structure at higher pH. While the melting curve for the reference Dickerson dodecamer showed the same sharp transition in all experiments, the transition for the [6X,7k]-dodecamer at high pH disappeared almost completely. Upon lowering the pH back to neutral, the original shape and higher melting temperature were recovered, indicating that the process is reversible and not the result of substitution or degradation of the oligonucleotide.

Table 2-4. pH-dependency of the melting temperatures for the reference Dickerson dodecamer and its [6X,7k] analog.

pH	Reference DiDo	[6X,7k] DiDo
7.0*	66.5 (0.7)*	55.5 (1.0)*
7.5	66.4 (0.9)	50.2 (1.6)
8.0	65.4 (0.6)	48.2 (2.0)
7.0	67.2 (0.7)	56.4 (0.9)

Both sequences were diluted in Breslauer buffer (10 mM sodium phosphate, pH 7.5, 1 mM EDTA, 1 M NaCl). Following the melting experiments, the pH was raised to 8.0 (NaOH, 0.1 M). The reversibility of the process was proven by lowering the pH back to 7.0 (HCl, 1M). Sample dilution in either case was negligible. *Reference data from previous high-salt experiments.

We conclude that the observed effect of pH on the melting behavior is largely the result of deprotonation of pyDAD. The gradual destabilization of the duplex structure, indicated by the decrease in the melting temperature was interpreted as a sign of increasing charge repulsion between the negatively charged puADA residues in the helix. In the absence of the positive charge on pyDAD, the formation of duplex becomes energetically unfavorable.

Salt-dependence of melting temperature. Besides studying the stabilizing effect of the pyDAD/puADA base pair in duplex nucleic acid, some of its structural features can be explored by melting studies. The self-complementary nature of the Dickerson dodecamer enables us to study the same sequence in two different conformations. At low concentrations of oligonucleotide and buffer salt, the intramolecular hairpin structure can be studied next to the regular double-stranded version. While pyDAD and puADA form a regular Watson-Crick base pair in duplex DNA, the hairpin formation places the nonstandard nucleotides in the single-stranded loop region (Figure 2-5).

Melting studies of the [6X,7k] dodecamer and the reference Dickerson dodecamer were conducted over a NaCl concentration range from 1 to 1000 mM. At low NaCl concentrations,

the expected biphasic-melting curve was observed (Figure 2-6). While the first transition was the result of duplex melting, the second step in the melting curve was assigned to the melting of the hairpin structure (Marky et al., 1983). Interestingly, duplex separation at the lower temperature was characterized by a sharp transition typical for strong cooperative behavior,

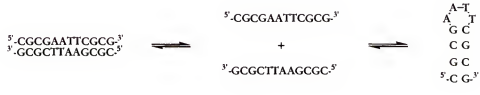
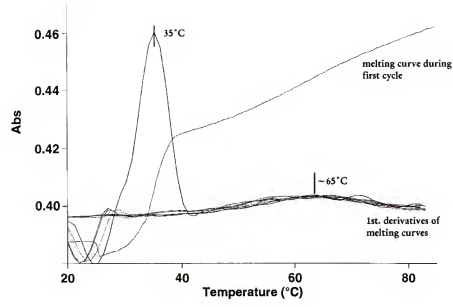


Figure 2-5. The Dickerson dodecamer exists in two possible conformations, because of its self-complementary nature. The two structures can be distinguished by the different melting temperatures. The duplex-to-single-strand transition occurs at lower temperature and is salt-dependent. The hairpin-to-single strand transition is the high temperature transition and not much affected by salt. Placing nonstandard nucleotides in the AT region of the oligonucleotide positions the base either opposite its hydrogen bonding partner (duplex DNA) or in the single-stranded loop region of the hairpin.

and was identical at 35°C for both sequences. Unfortunately, this transition was observed only during the initial heating cycle but was absent in all subsequent cycles, due to the unfavorable entropic factor of low salt concentration. In contrast, the second transition was broad, indicating multiple intermediates in melting of the hairpin, and was 5°C higher for the [6X,7k]-DiDo than for the regular dodecamer (Figure 2-6).

Increasing the NaCl concentration stepwise (Figure 2-7), the low-temperature transition reappeared at 100 mM NaCl. From there on, the melting temperature for the duplex increased steadily with increasing amounts of buffer salt. This pattern was observed for both, [6X,7k]-DiDo and reference DiDo, and reached steady state around 1 M NaCl. The trend was explained by reduced charge repulsion within the duplex with increasing salt concentration,

A



B

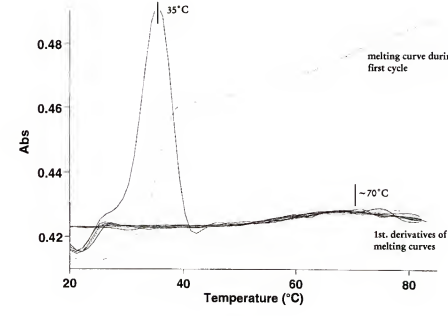
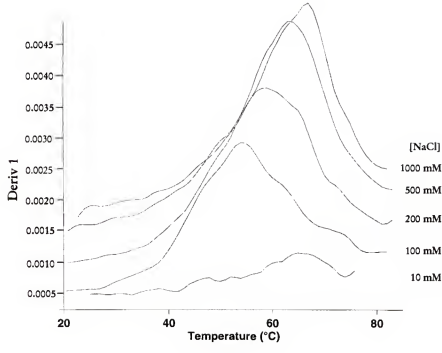


Figure 2-6. Regular melting curve overlaid by 1st derivatives of the reference Dickerson dodecamer (A) and [6X,7k]-DiDo (B) at 1 mM NaCl in the Breslauer buffer. Both melting curves show a sharp transition at 35°C for the duplex, followed by a broad transition around 60-70°C for hairpin denaturation (Marky et al., 1983). The total oligonucleotide concentration was 5 μM. Under these experimental conditions; the melting of the duplex around 35°C was observable in the first cycle. All consecutive rounds showed only the hairpin transition. A heating rate of 0.5-1°C/min, as well as the low oligonucleotide and salt concentration did not allow complete reequilibration between duplex and hairpin within a reasonable time period.

A



B

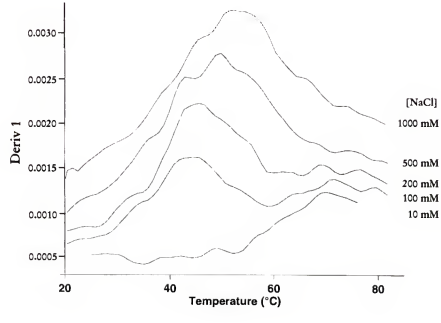


Figure 2-7. Salt-dependence of the melting temperature for the reference Dickerson dodecamer (A) and [6X,7k]-DiDo (B). The differentiated melting curves in Breslauer buffer at the indicated NaCl concentration show two transitions, the salt-dependent duplex-to-single-strand at lower temperature and the salt-independent hairpin-to-single-strand transition at high temperatures.

stabilizing the dimeric structure. The salt concentration around the duplex DNA structure becomes constant at 1.5 to 2 M NaCl, resulting in a constant T_M. As the duplex curve becomes more dominant and shifts towards higher melting temperatures, the second transition, the break up of the hairpin structure, turns into a shoulder and at concentrations above 200 mM NaCl, disappears as it is completely overlapped by the duplex transition. (Figure 2-7).

The melting temperatures of the two oligonucleotides show some interesting variations. Although the overall patterns of salt-dependence for both sequences are identical, the duplex transition for the nonstandard base pair-containing oligonucleotide is on average 10°C lower than the reference dodecamer. At the same time, the denaturing temperature for the hairpin structure that contains pyDAD and puADA is about 5°C higher.

Discussion

Based on the melting studies presented here, we conclude that pyDAD and puADA can form a base pair and contribute to the overall stability of duplex DNA. The observed 10°C-drop in the melting temperature for the sequences that contain nonstandard base pairs is significant but still higher than the 30-35°C reduction in melting temperature, determined for the same sequence that carries a GT wobble-base pair in that position. The same salt-dependence of the duplex transition, observed for the reference and the pyDAD/puADA oligonucleotide is a good indication of the regular dimeric behavior of the sequences with nonstandard base pairs.

Beyond the possibility that the temperature difference for pyDAD/puADA oligonucleotides is within the standard deviation of stabilizing and destabilizing effects as a result of sequence context (Kawase et al., 1986) (Table 2-3), the disruption of the minor-groove spine hydration could explain the observed pattern. In their analysis of the X-ray structure of

the Dickerson dodecamer, Drew and Dickerson (Drew & Dickerson, 1981) reported an unusually high degree of ordered water molecules in the -AATT- minor groove region, forming a spine of water. The additional hydrogen bonds in the minor groove of AT base pairs contribute to the overall stabilization of B-form DNA.

The pyDAD/puADA base pair could disrupt the formation of the stabilizing hydrate hull. Precedence for disruption of the minor-groove spine hydration by the NH_2-groups of guanosine can be found in CG-rich regions of nucleic acid (Conner et al., 1982).

More evidence for the significance of spine hydration in DNA structure comes from recent studies with 2-aminoadenosine (Bailly & Waring, 1998; Coll et al., 19896; Howard et al., 1984). Replacing adenosine by the 2-amino analog, the effect of the NH_2 group on the hydration structure could be measured. The results clearly indicated a higher tendency for a transition from the B to Z-form DNA.

How is this relevant to our oligonucleotides with the pyDAD/puADA base pairs? In our modified Dickerson dodecamer, the nonstandard base pair is placed in the center of the oligonucleotide, replacing one adenosine and one thymidine. As with guanosine and diaminopurine, the amino group of pyDAD could interfere with the formation of the minor groove spine hydration, causing configurational changes, which affect the melting temperature of our sequences. The noticeably broader transition for the [6X,7k]-DiDo (Figure 2-8) supports structural changes as a result of the disruption of minor-groove spine hydration remains unanswered. Finally, this interpretation would also explain the melting temperatures depression for an oligonucleotide, containing a pyDAD/puADA (π) base pair, reported by Piccirilli (Piccirilli et al., 1990).

The possibility of the syn-conformation of pyDAD in the context of an oligonucleotide could be responsible for the reduction in melting temperature. The actual

conformation around the C-glycosidic bond of pyDAD in an oligonucleotide is not known. Examples for syn-conformations in base pairs had been reported for purine bases (Kawase et al., 1989). In contrast to regular N-nucleosides, which can not form base pairs in such a structural arrangement, the syn-conformation of pyDAD still presents a pyDA pattern. The missing third hydrogen bond would then explain the lower melting temperature. At the same time, a pyDA hydrogen bonding pattern could compete with cytidine during primer extension experiments, resulting in misincorporation of guanosine opposite pyDAD in the template. No such misincorporation was observed, however (see Chapter 3).

Finally, the pH-dependency of the pyDAD/puADA sequence indicates the presence of stabilizing charges at pH 7. Upon removal of the partial protonation of pyDAD at pH >7.5, the duplex character is lost, increasing the destabilization of the structure due to the negative charges from the deprotonated puADA's.

The increased stability of the hairpin structure over the duplex, observed at low salt concentrations at pH 7 (Figure 2-6) could be explained by formation of additional hydrogen bonds with the surrounding water. The charged residues, all located in the loop region of the hairpin, could induce extensive structural arrangements, forming a multi-layer water hull. Such a hydrogen bonding network could significantly contribute to the overall stability of the monomeric structure and explain the increase melting temperature.

Conclusions

The building blocks for the pyDAD/puADA base pair were prepared by organic synthesis. The existing procedure for 2,4-diamino-2'-deoxy-5-ribofuranosyl-pyrimidine

(pyDAD) was optimized, doubling the yields of the 13-step synthesis. In addition, the correct β-conformation of the prepared 2'-deoxynucleoside was confirmed by NOE experiments. Furthermore, the stability of pyDAD under conditions used for oligonucleotide deprotection was investigated. No degradation or substitution of the pyrimidine nucleoside was observed.

The melting studies of a series of dodecamers have established the base-pairing ability of pyDAD and puADA. Nevertheless, melting studies are not conclusive in all aspects of the pyDAD/puADA base pair, making it impossible to draw final conclusions. Instead, we are left with many speculations about the significance of various effects that have been reported in the literature. Further insight into how pyDAD and puADA pair within an oligonucleotide duplex may come only from a crystal structure of the Dickerson dodecamer containing pyDAD and puADA. This work is currently in progress by our collaborators at UCLA.

CHAPTER 3
INCORPORATION OF NONSTANDARD NUCLEOTIDES BY HIV-1 REVERSE TRANSCRIPTASE

Introduction

The highest priority for developing further an artificial genetic system incorporating an expanded genetic alphabet is the identification of enzymes that recognize and incorporate nonstandard base pairs with sufficient specificity. Although the novel building blocks can be incorporated into oligodeoxyribonucleotides by chemical synthesis, in vitro selection and evolution experiments require enzymes such as kinases and polymerases that handle nonstandard nucleotides. Therefore, to study the significance of the expanded genetic alphabet in a biological context, we not only need to develop synthetic approaches for the unnatural nucleotides, but we also must explore, expand, and adjust the toolbox of molecular biology. New biomolecular tools, enzymes with broader or novel substrate specificity, must either be found or generated.

Extensive screening experiments of DNA polymerases and reverse transcriptases have indicated that most enzymes terminate primer elongation either at the position of the nonstandard nucleotide in the template, or immediately thereafter (Angerer & Ankenbauer, 1995; Horlacher et al., 1995; Lutz et al., 1996). Various mutants of HIV type-1 reverse transcriptase (HIV-RT) and, to a lesser extent, 9°N-7 (exo^{DID}), a version of 9°N-7 DNA polymerase with reduced exonuclease activity (Lutz et al., 1998; Southworth et al., 1996)

proved to be the best at incorporating the pyDAD-puADA base pair. Based on the work of Horlacher (Horlacher, 1995) and Lutz (Lutz, 1997), a selection of HIV-RT mutants, AZT-21, Y181, Y188L, and M184V, were studied in more detail.

HIV Type-1 Reverse Transcriptase

The reverse transcriptase from HIV type-1 is a heterodimer, built from a 66-kDa (p66) and a 51-kDa (p51) subunit. The p51 arises naturally from proteolytic cleavage of the C-terminus of p66. The enzyme possesses three distinct catalytic activities: RNA-dependent DNA polymerase, RNase H, and DNA-dependent DNA polymerase. It naturally catalyzes the reverse transcription of the viral RNA genome of the HIV into double stranded DNA, which is then integrated into the host genome, eventually leading to viral reproduction. The overall structure of HIV-RT resembles the typical "right hand" finger-palm-thumb arrangement found in all known DNA polymerases (Joyce & Steitz, 1995). The active site, located in the palm domain, carries the highly conserved aspartic acid residues (D110, D185, D186), which participate in the well known two metal ion mechanism (Steitz, 1998). (Figure 3-1)

In the fight against HIV infection and for a therapeutic treatment of AIDS, the uniqueness of the reverse transcriptase to the virus makes this enzyme an ideal target. To interfere with HIV replication, nucleosidic inhibitors (NRTI) such as 3'-azido-3'-deoxythymidine (AZT) and 2',3'-dideoxycytidine (ddC), and non-nucleosidic reverse transcriptase inhibitors (NNRTI) such as nevirapine or tetrahydrobenzodiazepine (TIBO) were identified. The susceptibility of HIV-RT to NRTIs may reflect the reverse transcriptase's high base substitution error frequency ($\sim 2 \times 10^{-4}$; (Kati et al., 1992)). The low fidelity of the enzyme presumably allows incorporation of the analogs, causing termination of replication upon incorporation.

For NNRTIs, kinetic studies established that inhibition was non-competitive versus dNTP and template by nature. Backed by the X-ray crystal structure of an HIV-RT-inhibitor complex (Kohlstaedt et al., 1992), Johnson and coworkers (Spence et al., 1995) proposed a common binding site for many NNRTIs on the backside of the active site, causing a dislocation of the highly conserved carboxylate side-chains D185 and D186, which are involved in metal ion ligation. The concentration of the NNRTI influenced the K_D for magnesium, as expected based on this

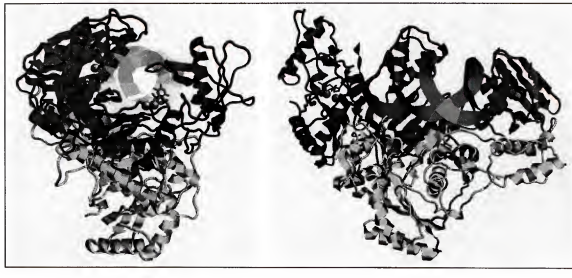


Figure 3-1. Front and side-view of an HIV type-1 reverse transcriptase complex, co-crystallized with primer-template and dNTP (Huang et al., 1998) (1RTD). The double stranded DNA (shown as ribbon) interacts exclusively with the p66 subunit (dark) of HIV-RT while p51 contributes to the structural integrity of the enzyme. Looking down the DNA helix, the front view nicely shows the finger (on the right side) and thumb domains (on the left) wrapping around the nucleic acid substrate. The active site is marked by the two magnesium ion spheres and the thymidine triphosphate in the palm region. The view from the side shows the active site of the polymerase to the left. The 5'-end of the primer is perfectly aligned to the incoming triphosphate while the single stranded template takes a 90° turn, protrudes on top of the polymerase. Similar structural features were identified for tertiary complexes of polymerase β (Pelletier et al., 1994). The RNase H domain is located on the right hand end of the p66 subunit. One magnesium ion was identified in the active site. Graphical layout were produced with RasMol 2.6 (RasMac Molecular Graphics).

mechanistic proposal. Alternatively, the inhibition may arise because the drug interferes with the thumb movement, a conformational change required for catalysis (Kopp et al., 1991; Tantillo et al., 1994).

Additional evidence for the location of the proposed binding sites for NRTIs and NNRTIs comes from structural studies of HIV-RT mutants that are resistant to these drugs. Within a relatively short time period following introduction of NRTIs and NNRTIs for AIDS therapy, resistant strains of HIV emerged. Sequence analysis of these mutants show distinct patterns and regional preferences for modifications, depending upon the drug applied. Resistance to NRTIs is achieved mainly by modifications in the dNTP and primer/template binding regions in the finger domain of the polymerase (Kerr & Anderson, 1997; Ueno & Mitsuya, 1997). Evolution in the presence of NNRTIs results in mutations in an area on the backside of the active site in the palm domain (Fan et al., 1995; Smerdon et al., 1994).

HIV reverse transcriptase mutants and nonstandard base pairs

The chance of identifying a reverse transcriptase mutant from HIV that accepts additional letters in an expanded genetic alphabet are higher than these for other DNA polymerases for two reasons. First, HIV-RT has by nature a high tolerance for structurally modified nucleotides. In contrast to cellular polymerases, its substrate specificity towards nucleoside triphosphates is significantly lower. In addition, HIV-RT possesses no exonuclease mechanism to remove nucleotides in a stalled complex. Second, the diversity of naturally occurring mutant forms of HIV-RT may indicate a high flexibility of the enzyme to adapt to variations in substrates without losing its catalytic properties. Evolution of resistance by HIV-RT against specific drugs has generated polymerases that often behave differently with unnatural substrates other than the one for which they were selected for. An example for this

behavior is the increased sensitivity of 3TC-resistant HIV-RT to AZT and *vice versa*, which has been successfully applied in AIDS therapy to lower the viral load (Larder et al., 1995).

Encouraged by the prescreening results reported by Horlacher (Horlacher et al., 1995) and Lutz (Lutz, 1997), we more closely examined the four HIV-RT mutants AZT-21, M184V, Y181I, and Y188L that showed some specificity for an additional base.

HIV reverse transcriptase mutant AZT-21 and M184V

HIV-RT mutants AZT-21 and M184V both emerge from patients treated with NRTIs. The AZT-21 mutant carries five mutations, and confers resistance to AZT (Kellam et al., 1992; Larder et al., 1989; Larder & Kemp, 1989). Mutations M41L, D67N, K70R, T215Y, and K219Q are all located in the finger domain (Figure 3-2). The discussion for a mechanism by which these mutations increase the discrimination against AZT is still ongoing. On one side, recent X-ray crystallographic studies on mutant enzyme propose that residues 41, 67, and 70 are involved in DNA template alignment, while the side chains of 215 and 219 induce conformational changes in the active site, similar to the effect proposed for NNRTIs (Ren et al., 1998). On the other hand, the arrangement of the mutations in the finger domain, thought to be the binding site for the incoming nucleoside triphosphate, may also allow for some direct interactions that affect fidelity (Boyer et al., 1994; Tantillo et al., 1994).

An interesting observation, unnoticed so far, is that aspartate 67 and lysine 219 form a salt bridge in native HIV-RT. The mutation of position 219 maybe only indirectly related to AZT resistance. Instead, the mutation of K219, which has no measurable effect on AZT sensitivity by itself (Larder et al., 1991), may be an example of charge compensation after mutagenesis of the bridging partner to achieve AZT resistance. This could easily explain the difficulties in directly correlating all of the observed mutations with AZT insensitivity.

The replacement of methionine 184 by an alanine or a valine arises in patients treated 2',3'-dideoxyinosine (ddI). In addition to discrimination against dideoxy compounds, M184 mutants show very high specificity against misincorporation of 3'-thia-2',3'-dideoxycytidine (3TC) and 5-fluoro-3'-thia-2',3'-dideoxycytidine (FTC). Located in the palm domain, next to the highly conserved two aspartic acids of the active site (Figure 3-2), the mutation leads to an increased fidelity of the HIV-RT (Bakhanashvili et al., 1996; Hsu et al., 1997).

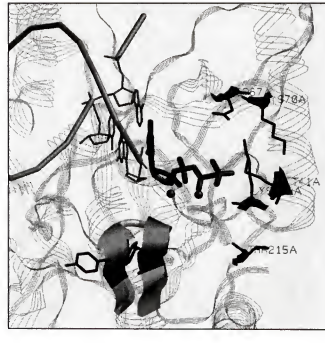


Figure 3-2. The active site of the HIV-RT, co-crystallized with primer-template and TTP (Huang et al., 1998) (1RTD). While the template (shown in backbone) takes a sharp 90° turn up, the 3'-end of the primer is positioned next to the incoming TTP. The crystal structure suggests proper orientation of the base through hydrogen bonds with the adenosine in the template while the triphosphate portion of TTP is locked in position by the two magnesium ions, represented by the spheres. These, in turns, are held in position by two aspartic acids, located in the loop region of an antiparallel β-strand on the bottom of the picture. Y181 and Y188 are part of this β-strand, indicated by the wireframe structures. The mutations, resulting from AZT treatment, are shown on the right hand side.

Mechanistic and structural studies on the effects of M184 mutations place the amino acid side chain of position 184 near the deoxyribose of the primer's 3'-end or the bound dNTP, suggesting that the residue participates in nucleotide substrate binding (Pandey et al., 1996; Tantillo et al., 1994; Wilson et al., 1996).

HIV reverse transcriptase mutant Y181I and Y188L

In patients treated with the NNRTIs nevirapine and TIBOs, HIV type-1 reverse transcriptase responds with several characteristic mutations in the lower palm region (Smerdon et al., 1994) (Figure 3-2). The most common mutation replaces tyrosine 181 by cysteine. Position 181, together with the tyrosine 188 are thought the two key residues for NNRTI interactions, based on photoaffinity labelling with a nevirapine analog (Cohen et al., 1991). The binding is attributed largely to π-π stacking interactions between the tyrosines and the aromatic groups of the inhibitors. The Y181C mutation alters the shape of the binding pocket, reducing the affinity of the NNRTI (Smerdon et al., 1994). A similar effect can be achieved by substitution of the two positions with the amino acids, found in the same location in HIV type-2 RT. The closely related variant carries an isoleucine at position 181 and a leucine at 188, and shows no inhibition in the presence of NNRTIs. Introducing either of the two point mutations into HIV type-1 RT desensitizes the enzyme against the inhibitors (Sardana et al., 1992). Simultaneous mutation of both positions improves the resistance qualitatively and quantitatively (Fan et al., 1995). The mechanism of resistance is therefore a simple modification of the enzyme's surface, critical to NNRTI binding. How and to what extent these tyrosine substitutions effect the polymerase's performance are not very well understood (see also Conclusion).

Material and Methods

Experimental Procedure

Reagents for DNA synthesis were purchased from Glen Research (Sterling, VA). Regular oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA) and purified by PAGE (Appendix C8/C10). Oligonucleotides containing nonstandard nucleotides were prepared as described in Chapter 2 - Material and Methods.

Protein expression, purification, and activity measurement. The plasmid DNA of the various mutants of HIV type-1 reverse transcriptase were a gift by S.H. Hughes and P.L. Boyer (NCI - Frederick Cancer Research and Development Center, Frederick, MA) (Boyer et al., 1994). Expression and purification of the wild-type HIV-RT, as well as the four mutants Y181I, M184V, Y188L, and AZT-21, was done by Jeong Ho Park.

Enzyme activity was determined by incorporation of [^{3}H]-TTP into a poly(rA)-oligo(dT) template (Bryant *et al.*, 1983; Kuchta, 1996; Reardon & Miller, 1990; Stahlhut & Olsen, 1996). All reactions were carried out at 37°C in a water bath. An aliquot of the polymerase preparation (1–2 μL) was incubated in HIV-RT buffer (3x) in the presence of 5 μg poly(rA)-oligo(dT)$_{12\text{-}18}$ (Pharmacia) and [^{3}H]-TTP (25 μM, 6000 cpm/pmol, Amersham; concentration adjusted with TTP (1 mM)). Four aliquots (20 μL) were taken over a 12 min period and quenched with EDTA (10 μL, 0.5 M, pH 8). Twenty microliters of the quenched reaction mixture were applied to 2.5-cm circles of Whatman DE-81 filter paper. The air-dried filter papers were washed three times with disodium phosphate (0.15 M), twice with ethanol, and finally once with diethylether. The dry filters were counted by liquid scintillation counting in 5 mL ScintiSafe 30% (Fisher Sci.). All experiments were repeated three times and the resulting data averaged. The activity was calculated from the slope of a time vs. cpm plot

and was expressed as units per mL (U/mL). One unit of enzyme was defined as the volume of polymerase preparation that converts 1 nmol TTP into filter-bound material in 10 min at 37°C.

Oligonucleotide synthesis: primers and templates. The following sequences were synthesized for the primer extension experiments and PCR amplifications. The reaction conditions were adjusted for the nonstandard nucleotides as described previously (Chapter 2). k marks the position of pyDAD and X the position of puADA.

P3-1: 5'-GCG AAT TAA CCC TCA CTA AAG-3' (T3 primer sequence)

P3-2: 5'-GCG TAA TAC GAC TCA CTA TAG-3' (T7 primer sequence)

P3-3: 5'-GCG AAT TAA CCC TCA CTA AAG AAC G-3'

P3-4: 5'-GCG TAA TAC GAC TCA CTA TAG ACG A-3'

P3-5: 5'-ACA GXA CTX GGX TT-3'

P3-6: 5'-GCG AAT TAA CCC TCA CTA AAG TAC G-3'

T3-1: 5'-GCGAATTAACCCTCACTAAAGTACGTTCGTCTATAGTGAGTCGTATTACGC-3'

T3-2: 5'-GCGTAATACGACTCACTATAGACGTTCGTTCTTTAGTGAGGGTTAATTCGC-3'

T3-3: 5'-GCGAATTAACCCTCACTAAAGTACGkTCGTCTATAGTGAGTCGTATTACGC-3'

T3-4: 5'-GCGTAATACGACTCACTATAGACGTXCGTTCTTTAGTGAGGGTTAATTCGC-3'

Primer extension experiments. In a typical primer extension experiment, primer (105 pmol; 5'-^{32}P-labelled with polynucleotide kinase – see Appendix C7) and template (147 pmol) in HIV RT buffer (56 μL, 3x) were mixed with dATP, dGTP, dCTP, and TTP (final concentration: 130 μM each) and the total volume adjusted to 160 μL with water. In experiments with nonstandard nucleotides, the concentrations for d(pyDAD)TP or

d(puADA)TP was the same as those for standard dNTPs, if not mentioned otherwise. The additional volume was compensated by the amount of water added to the reaction. After heating the sample to 95°C for 1 min, the primer/template complex was annealed by cooling the sample to room temperature over 1h. The primer extension was started by addition of the polymerase (16 μL) and the reaction mixture incubated at 37°C. Aliquots (25 μL), taken at various times during the reaction, were quenched by addition of a premixed solution of sodium acetate (2.5 μL, 3 M, pH 5.2), EDTA (1 μL, 0.5 M, pH 8), and ethanol (50 μL). After storing the sample at -20°C for 20 min, the samples were centrifuged (14,000 rpm, 4°C, 20 min) and the pellets dried in the vacuum concentrator. The residues were redissolved in PAGE loading buffer and the samples separated on a 10% PAGE gel (7 M urea). The gel was analyzed using the MolecularImager®.

To improve reproducibility in cases where multiple reactions were run in parallel, a master mixture of primer/template and the dNTPs was prepared by simply scaling up the listed procedure. Master mixtures may not be stored for more than 24 hours at -20°C.

Postsynthetic 5'-biotinylation of oligonucleotides. A method to postsynthetically introduce a biotin residue onto the 5'-OH of oligonucleotides was developed. Following the basic protocol of a commercially available labelling kit (FluoroAmp™T4 Kinase Biotin Oligonucleotide Labelling System from Promega, Madison WI), the free 5'-OH of the oligonucleotide (~3 μmol) was thiophosphorylated with γ-S ATP by T4 polynucleotide kinase. After 2h at 37°C, the reaction was quenched by ethanol precipitation. The pellet was dried in the vacuum concentrator before being redissolved in PBSE-buffer (12.5 μL, supplied with labelling kit). For the actual biotinylation, the kit's biotin maleimide reagent was replacing by a solution of N-(biotinoyl)-N'-(iodoacetyl)-ethyldiamine (1.27 mg, Molecular Probes, Eugene OR) in DMSO (55 μL, anhyr). The biotinylation reagent (5 μL) was added to

the oligonucleotide solution and the mixture was incubated for 1 h at 68°C in the dark. Again, the reaction was quenched by ethanol precipitation. The dried pellet was purified by gel-filtration (Sephadex® G-50, supplied with labelling kit). The product fractions were lyophilized, redissolved in 100 μL water and the DNA concentration determined by UV absorbance. In addition, the quality of the reaction was checked by anion-exchange HPLC (Macrosphere 300A WAX 7U (Alltech, Deerfield IL) 4.6 mm x 250 mm; solvent A = $KHPO_4$ (20 mM, pH 6); solvent B = $KHPO_4$ (20 mM, pH 6), KCl (2 M); solvent C = acetonitrile; linear gradient from A/B/C (90/0/10) to (40/50/10) in 50 min).

Second round primer extension. Primer P3-2 (0.5 μmol) was mixed with 5'-biotinylated template T3-3 (0.6 μmol, see procedure above) and HIV RT buffer (266 μL, 3x). dATP, dCTP, dGTP, TTP, and d(puADA)TP (final conc.: 130 μM each) were added and the reaction volume was adjusted to 800 μL with water. Following primer annealing (95°C, 3 min and cooling to RT over 1 h) and addition of HIV-RT Y188L (3 U, 64 U/mL), the mixture was incubated at 37°C for 24 h. The reaction was quenched with EDTA (10 μL, 0.5 M, pH 8), loaded onto a preequilibrated streptavidin-agarose column (0.6 mL bed volume, Fluka Chemicals), washed with 6 volumes of Tris-HCl (10 mM, pH 7.5), NaCl (50 mM) , EDTA (1 mM, pH 8), and the unbiotinylated strand eluted with 5 volumes 0.2 M NaOH. The elongated primer product P3-2* was detected by UV absorbance at 260 nm and all product fractions were neutralized with sodium acetate (100 μL, 3 M, pH 5.2). The DNA was recovered by ethanol-precipitation and the pellet was redissolved to 2 pmol/μL.

For the primer extension in the opposite direction (second round), ^{32}P-labelled primer P3-1 (7.5 pmol) and template P3-2* (10 pmol) in HIV RT buffer (5 μL, 3x) were annealed after heat denaturation. To the primer/template complex, the standard dNTPs and d(pyDAD)TP (final conc.: 130 μM each) were added and the volume adjusted to 15 μL with water. Following

addition of HIV-RT Y188L (0.1 U, 64 U/mL), the mixture was incubated at 37°C. Sample aliquots (3 μL), taken from the reaction mixture over time, were quenched and precipitated with a premixed solution of water (7 μL), sodium acetate (1 μL, 3 M, pH 5.2), EDTA (0.2 μL, 0.5 M, pH 8), and ethanol (25 μL). The dried pellets were redissolved in 3.5 μL PAGE loading buffer and separated on a 10% PAGE gel (7 M urea). The gel was analyzed, using the MolecularImager®.

Standing start experiments. Primer (15 pmol, 5'-^{32}P-labelled – Appendix C7) and template (21 pmol) were mixed with HIV RT buffer (8 μL, 3x) and the reaction volume was adjusted with water to 21 μL. The samples were heat-denatured (95°C, 1 min) and cooled to room temperature over one hour. After addition of the appropriate dNTP (1.67 μL, final conc.: 130 μM) and a polymerase aliquot, the reaction was incubated for up to 30 min at 37°C. The reaction was quenched by addition of a premixed solution of sodium acetate (2.5 μL, 3 M, pH 5.2), EDTA (1 μL, 0.5 M, pH 8), and ethanol (50 μL). After ethanol precipitation, the pellet was dried in the vacuum concentrator. The residue was dissolved in PAGE loading buffer before loaded on a 10% PAGE gel (7 M urea). The gel was analyzed with the MolecularImager®.

PCR amplification experiments. A typical PCR experiment run over 5 to 10 cycles. The total volume of the reaction was 100 μL, containing primer P3-1 and ^{32}P-labelled P3-2 (1 μM each), as well as template T3-3 (20 nM) (or T3-1 for reference experiments), HIV-RT buffer (33 μL, 3x, pH 8.0), and all the dNTPs (200 μM each). The volume was adjusted with water. The primer were annealed by heating the PCR tube to 95°C for 1 min, followed by letting it cool to 37°C over 10 min. After each denature-annealing round, a fresh batch of HIV-RT mutant Y188L (0.25 U, 64 U/mL) was added and the reaction mixture incubated at 37°C. The elongation time in the reference experiments with standard nucleotides and HIV-RT was 1 h at

37°C. In the presence of nonstandard nucleotides, the reaction time was extended to 24 h. Aliquots for PAGE analysis (10 μL) were taken after reannealing, prior to addition of fresh enzyme. The samples were quenched with a premixed solution of sodium acetate (1 μL, 3 M, pH 5.2), EDTA (0.2 μL, 0.5 M, pH 8), and ethanol (25 μL). After ethanol precipitation, the dried pellets were resuspended in PAGE loading buffer and separated on a 10% PAGE gel (7 M urea). The gel was analyzed with the MolecularImager®.

DNA sequencing with puADA. Deprotected and PAGE-purified oligodeoxyribonucleotide P3-5 (0.1 OD, 5'-^{32}P-labelled – Appendix C7) was split into three equal portions (8 μL each). While the first aliquot was kept for reference purposes, the second was incubated with formic acid (3 μL, 8.8%) for 7 min at 37°C. The reaction was quenched with piperidine (150 μL, 1 M), followed by strand cleavage at 90°C (30 min). The third aliquot was directly incubated with piperidine (100 μL, 1 M) at 90°C (30 min). After lyophilization and co-evaporation with water to remove the excess piperidine, all samples were dissolved in 10 μL PAGE loading buffer and separated on a 20% PAGE gel. The gel was analyzed with the MolecularImager®.

Results and Discussion

Incorporation of pyDAD and puADA by HIV Reverse Transcriptase

A selection of HIV type-1 reverse transcriptase mutants (Y181L, Y188L, M184V, and AZT-21), previously identified as possible candidates for PCR-like amplification in the presence of pyDAD and puADA, were analyzed by primer extension and standing start experiments.

Primer extension experiments

The primary investigation of the four mutants of HIV-RT focused on the ability of the template-directed primer extension. A single nonstandard nucleotide (pyDAD) was placed at position 25 in the template (T3-3), four bases upstream of the 5'-end of the primer. The template was annealed with the radiolabelled primer (P3-2), the complex was split into aliquots, and each aliquot was separately incubated with polymerase in two tubes, one containing only the standard nucleoside triphosphates, the other containing the regular dNTPs plus d(puADA)TP. At time intervals, samples were removed from the reaction mixture and the progress of the primer extension reaction was followed by PAGE as shown in Figure 3-3.

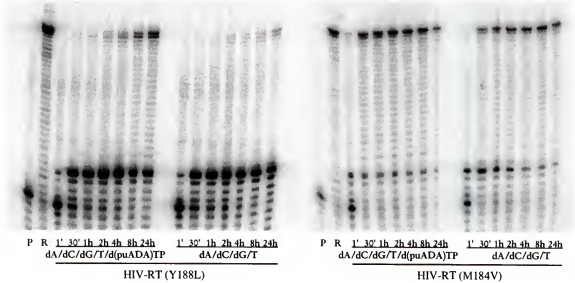


Figure 3-3. Primer extension experiments with HIV-RT mutants, using P3-2 (21-mer) and T3-3 (51-mer). HIV-RT mutants Y188L (15 U/mL) and Y181I (95 U/mL; data not shown) both produced more full-length product in the presence of d(puADA)TP than in its absence. For M184V (78 U/mL) and AZT-21 (177 U/mL; data not shown), the band pattern is almost undistinguishable, indicating low specificity for the nonstandard nucleotide. Experiments were performed at pH 7.2 (Materials and Methods). Gels were quantified by integration of the radioactivity in each lane and the individual bands. For comparison, the relative amounts were calculated from the ratio of counts for the specific band divided by the total counts per lane.

Based on their performance, Y188L and Y181I became the lead polymerases. Y188L and Y181I generated 4-5 times more full-length product in the presence of the standard nucleotides and d(puADA)TP than in the experiments without the d(puADA)TP over the same time period. The same experiments for HIV-RT wt, M184V, and AZT-21 showed no significant difference in the amount of full-length product synthesized in the presence or absence of d(puADA)TP. Standing start experiments with wt, M184V, and AZT-21, conducted later, confirmed the infidelity of these polymerases. Misincorporation of dATP opposite pyDAD and dCTP/TTP opposite puADA were seen the most. Furthermore, HIV-RT wt, M184V, and AZT-21 all showed very high tendency for G-T mismatches.

The two lead polymerases (Y188L, Y181I) were also tested in the opposite direction with primer P3-1 and the puADA-template T3-4. For both enzymes, the gel band patterns for the experiments with and without d(pyDAD)TP were indistinguishable, suggesting no specificity towards the nonstandard bases (data not shown). Standing start experiments, studying d(pyDAD)TP incorporation directly, proved the successful primer extension although with clear competition from dCTP and TTP.

In conclusion, the Y188L mutant of HIV-RT was identified as the most faithful polymerase for accepting the additional base pair. Although Y181I performs as well as Y188L in the primer extension experiments, the standing start experiments indicate a generally lower fidelity, represented by a strong G-T mismatch (data not shown). In pH-dependency studies, the d(puADA)TP incorporation opposite pyDAD by HIV-RT Y181I improved with decreasing pH. Unfortunately, the reduction in the pH of the reaction buffer below pH 7 lead to a drastic drop in primer extension in the opposite direction. Most likely, the protonation of d(pyDAD)TP (pK_A = 6.7) reduces its availability to the enzyme, resulting in an increase of misincorporation of one of the standard nucleotides.

Primer extension over two rounds

Based on the successful primer extension experiment by HIV-RT Y188L, primer extension in the reverse direction (incorporation of d(pyDAD)TP opposite puADA in the template), using the previously elongated primer as a template, was attempted. Early experiments indicated that the original template from the first round of primer extension must be removed. Otherwise, primer annealing for the second round was possible only in the presence of a large excess of radiolabelled primer, which in turn made the detection of the resulting small quantities of full-length product at the end of round two impossible. To separate the newly synthesized strand from the elongated primer/template complex, the original template was postsynthetically modified with a biotin moiety at the 5' end. Respective kits are commercially available and the only modification of the manufacturer's protocol was the substitution of the biotin-malemide reagent by the more base-stable N-(biotinoyl)-N'-(iodoacetyl)-ethyldiamine (see Materials and Methods). Following the first round primer extension, the reaction mixture could be loaded onto a streptavidin column. Upon treatment with sodium hydroxide, the double-strand falls apart and the newly synthesized complementary strand was eluted. The recovered fractions may contain some unreacted primer, as well as incompletely elongated material. Although these contaminants probably affect the UV readings, no interference in the following second round of primer extension was expected.

The second round primer extension by HIV-RT mutant Y188L in the presence of a ^{32}P-labelled primer is shown in Figure 3-4. To better visualize the difference in primer elongation, the relative amount of primer and full-length product were plotted versus time. Also shown in the plot is the amount of extended primer that accumulates as a result of pausing of the enzyme after position 25, the puADA in the template.

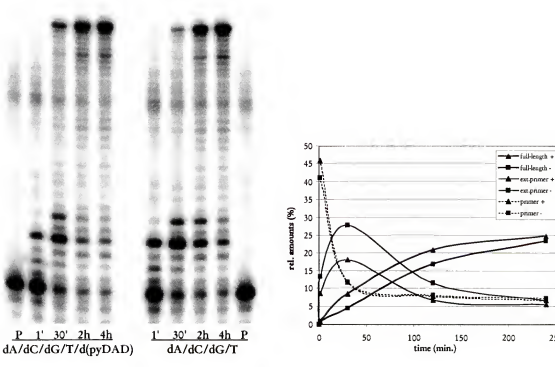


Figure 3-4. Second round primer extension in the presence or absence of d(pyDAD)TP. Samples were taken after the indicated incubation time. The relative amount of primer, extended primer, and full-length product were determined from the proportion of radioactivity in the DNA reaction product to the entire radioactivity in the lane. Plotting the relative amounts versus reaction time, the rapid decrease of primer, the temporary appearance of the extended primer and finally formation of full-length product for a primer extension in the presence and absence of d(pyDAD)TP can be monitored. Gels were quantified and the relative amounts were calculated as described in Figure 3-3.

Probably the best evidence for the presence of puADA in the 2nd round template is the observed pausing of HIV-RT at the predicted position. Additional proof was provided by an experiment with the same primer/template complex and *Taq* DNA polymerase, which resulted in 100% termination at position 25 (data not shown).

Of interest is also the second, weaker band that appears at position 27. In a recent publication, Giese and coworkers (Marx et al., 1999) reported the same phenomena in primer extension experiments with 4'-modified nucleotides and HIV-RT. They concluded that the pausing intermediates, 1-3 bases downstream of the 5'-end of the primer, were the result of

unfavorable interactions between the 4'-acetyl modification in the backbone of the primer/template complex and the enzyme in the DNA binding cleft.

Although the stalling for d(pyDAD)TP occurs in a different position, this additional band in the gel may provide evidence for structural distortion, introduced by the nonstandard nucleotide after incorporation in an oligonucleotide. Whether the stalling of primer elongation in the case of pyDAD is a result of direct steric interference of the primer with enzyme is not certain. Alternatively, the pausing could be indirectly caused by conformational changes in the primer/template complex, thereby preventing the 5'-end of the primer from aligning in a catalytically proper way in the active site. Either possibility must be considered until structural proof by NMR or X-ray crystallography is obtained.

A plot of the relative band intensities versus the reaction time revealed some more details about the course of the reaction. The disappearance of primer for both reactions is identical, suggesting that the initial polymerase activity is similar. The time course for build-up of extended primer is approximately the same for both reactions. The amount of stalled primer differs almost twofold, however. This observation is also reflected in the clearly faster formation of full-length product in the presence of d(pyDAD)TP.

In summary, the primer extension experiments over two rounds proved the specific incorporation of a d(puADA)TP opposite pyDAD in the template by the Y188L mutant of HIV-RT. Following partial purification of the newly synthesized template, primer elongation in the opposite direction, incorporating d(pyDAD)TP with some specificity vis-à-vis puADA in the template, was shown. Although the entire experiment was done with only one nonstandard base pair in the sequence, the results are encouraging for further investigation of HIV-RT in the search for a lead polymerase that can selectively handle an expanded genetic alphabet.

Standing start experiments

A series of single nucleotide primer extension experiments were performed with HIV-RT Y188L to identify the nucleotides most likely to compete with the nonstandard nucleotides during primer extension. In addition, the pH effect on the fidelity was addressed by these experiments. As for d(puADA)TP, HIV-RT mutant Y188L performed very well in standing start experiments. The incorporation of d(puADA)TP and all the standard nucleotides opposite pyDAD was investigated over a pH-range from 6.0 to 8.0. Not surprisingly, no primer extension was detected below pH 6.5. This finding agrees well with the pH-activity profile of HIV-RT, suggesting an inactive enzyme at low pH. Over the remaining pH-range, d(puADA)TP was incorporated with very good specificity vis-à-vis pyDAD (Figure 3-5). No significant misincorporation of any standard nucleotides was observed. The fact that puADA is mostly N3-deprotonated (pK_A of puADA: 5.7) under the reaction conditions does not seem to affect the enzyme's performance.

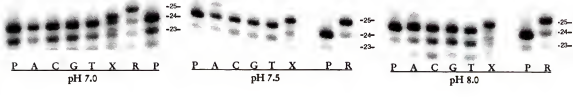


Figure 3-5. Standing start experiments for the incorporation of dNTPs opposite pyDAD. The experiments at pH 7.0, 7.5 and 8.0 are shown. d(puADA)TP is marked X. Radiolabelled primer P3-4 was annealed to the template T3-3 and incubated with 0.2 U of HIV-RT (Y188L) for 30 min at 37°C. Unreacted primer P3-4 (P) and a control reaction R (P3-4/T3-1 with dATP) were run in reference. The extended primer, carrying the puADA, shows a different migration pattern. The primer+1 band moves faster than the extended primer from the positive control while the mobility of the unreacted primers in unaffected. This variation in the presence of puADA was repeatably observed and is probably caused by additional negative charge as a result of deprotonation at N-3 of puADA (pK_A = 5.7) (Roy & Miles, 1983).

The standing start experiments in the other direction, studying the incorporation of d(pyDAD)TP, were less conclusive. Early results from chemically synthesized templates containing puADA identified TTP as the main competitor to d(pyDAD)TP for incorporation opposite puADA. These findings were not reproduced at a later stage using templates from newer DNA syntheses, as well as the enzymatically synthesized template P3-2* (see Primer extension over 2 rounds). Instead, dCTP becomes the new major competitor as seen in Figure 3-6. The results were unchanged when fresh dCTP was used, ruling out the possibility that dUTP, formed by deamination of dCTP, was responsible for the result.

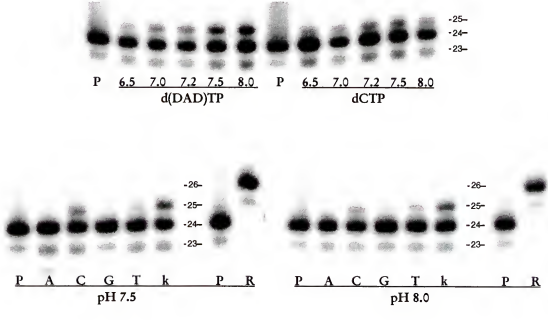


Figure 3-6. pH dependency in standing start experiments for the incorporation of dNTPs opposite puADA. The incorporation of d(pyDAD)TP shows much stronger pH dependence than d(puADA)TP in Figure 3-4. By increasing the pH of the reaction, the ratio of d(pyDAD)TP vs. dCTP incorporation shifts in favor of the nonstandard nucleotide. Standing start experiments at high pH conditions indicate a reasonable selectivity for pyDAD over dCTP. For all experiments, radiolabelled primer P3-3 was annealed to the template T3-4 and incubated with 0.2 U of HIV-RT (Y188L) for 30 min at 37°C. Unreacted primer P3-3 (P) and a control reaction R (P3-3/T3-2 with dATP) were run in reference.

As discussed previously, Eritja and coworkers (Eritja et al., 1986) had earlier reported that *Drosophila* DNA polymerase α preferred misincorporation of TTP, closely followed by dCTP, opposite puADA. Since the chemical synthesis of puADA-containing oligonucleotides has repeatedly caused problems (high termination ratio on the DNA synthesizer) and requires harsh deprotection conditions, it is likely that the observed misincorporation of TTP may be the result of partial transformation of puADA in the template rather than true mismatching by the polymerase. The results from the enzymatically prepared template strand are probably the most reliable.

The improved incorporation of d(pyDAD)TP at higher pH can be explained by the deprotonation of pyDAD ($pK_A = 6.7$). Around pH 7, almost half the triphosphate of pyDAD is protonated. Getting the positive charge in the active site of a polymerase could be highly unfavorable. In addition, the formation of hydrogen bonds would be restricted by the positive charge. As the pH of the reaction mixture increases, the fraction of protonated material decreases rapidly, making d(pyDAD)TP an acceptable substrate for a polymerase.

In summary, the standing start experiments have indicated the possibility for specific incorporation of nonstandard nucleotides opposite each other without major interference from standard nucleotides. We have shown that HIV-RT can both recognize and catalyze incorporation of the additional letters of the genetic alphabet. However, the question of what happens after catalysis remains open. As mentioned earlier, polymerases almost always pause after incorporation of the unnatural or functionalized nucleotide. Although we have provided evidence that HIV-RT incorporates the correct hydrogen bonding partner, a PCR experiment still fails if the elongation is terminated immediately afterwards. While extended reaction times

can partly solve the problem, another way to push the enzyme towards a more efficient primer extension would be to increase the polymerase concentration in the reaction mixture.

PCR Amplification with HIV Reverse Transcriptase

Equipped with a polymerase that incorporates puADA and pyDAD and the optimized reaction conditions that should maximize fidelity for the six letters of the genetic alphabet, a PCR-like amplification of oligonucleotides was attempted. A test of the experiment's design with the standard oligonucleotide template T3-1 is shown in Figure 3-7A. The gel shows an increasing amount of full-length product, verifying the possibility for PCR-like amplification with HIV-RT under the given conditions.

The PCR amplification with the nonstandard nucleotides was performed by using template T3-3. All conditions were identical to the reference experiments with the exception of an extended elongation period of one day to enable the reaction to go to completion. The analysis of the resulting PAGE gel indicated problems with degradation of the oligonucleotides under these conditions (Figure 3-7B). Although full-length product appears after the first two or three cycles, the entire reaction mixture is slowly digested. Since HIV-RT does not posses any DNA exonuclease activity, the enzyme preparation most likely contains some host contamination. However, no additional bands were visible in the SDS-PAGE of the enzyme solution (Jeong Ho Park – personal communication).

In a follow-up experiment, an identical reaction mixture without template was incubated for 3 days in the presence of HIV-RT Y188L. Samples were taken over the entire course of the reaction and analyzed by PAGE. Only very little degradation was observed. These data, as well as the stop of primer degradation in the latter cycles of the PCR (Figure 3-7B) may indicate that the exonuclease only works on double-stranded DNA. To overcome the primer digestion, an HIV-RT sample with a specific activity of up to 1500 U/mL that allowed

a reduction of the reaction time to 1 h was tested for amplification. Unfortunately, the exonuclease activity seems to co-concentrate with the enzyme, resulting in no improvement (data not shown).

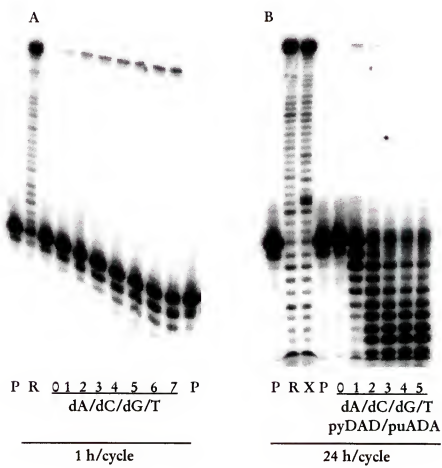


Figure 3-7. PCR-like amplification of oligonucleotides. A) testing the ability of HIV-RT Y188L to perform PCR-like amplification with standard template T3-1 and the regular dNTPs over 7 cycles, each 1 h (37°C). A plot of the relative amount of full-length product per lane shows a linear increase in product. B) PCR amplification of T3-3 with pyDAD and puADA over 5 cycles, each 24 h (37°C). Exonuclease contamination of the HIV-RT results in degradation of primer and full-length product that appears after the first two rounds. In both experiments, P marks the primer and R the positive control (primer extension of P3-2/T3-1 complex by *Taq* DNA polymerase). X marks the positive control for the first round primer extension (P3-2/T3-3 with HIV-RT).

Summarizing the results from the amplification experiments, we conclude that HIV-RT might be used for the PCR, provided that it is obtained in a form entirely free from nuclease activity. Although the low amounts of nuclease do not greatly hinder normal primer extension experiments, the degradation of oligonucleotides prevents PCR-like template amplification.

DNA Sequencing with Nonstandard Nucleotides

A theme relating to the expanded genetic alphabet concerns the development of new method to detect unnatural nucleotides in an oligonucleotide. Although very efficient methods for the identification of standard nucleotides exist, protocols for placing pyDAD and puADA in a sequence need to be developed. Enzymatic sequencing by Sanger dideoxy sequencing requires a) a functional polymerase and b) the 2',3'-dideoxy nucleoside triphosphates of all building blocks. While searching for a polymerase for pyDAD and puADA, the nonstandard dideoxy nucleotides can be synthesized from ddGTP (for puADA) and by a second deoxygenation of *N*2,*N*4-di-(t-butyl-benzoyl)-2,4-diamino-5-(5'-O-(4',4''-dimethoxy-trityl)-2'-deoxy-β-D-ribofuranosyl)-pyrimidine (**12**) (see Chapter 2) giving access to 2',3'-dideoxy(pyDAD) (data not shown).

Chemical sequencing, following the protocol by Maxam-Gilbert (Ambrose & Pless, 1987; Maxam & Gilbert, 1980; Sambrook & Manatis, 1983) is the alternative. The C-nucleosidic pyDAD, not cleavable under these conditions, can be detected by the absence of a degradation band in the sequencing gel ladder. To identify puADAs in an oligonucleotide sequence, very mild depurination conditions have shown to selectively cleave the xanthine-sugar bond without seriously affecting the other purine bases. The procedure was successfully

demonstrated on two oligonucleotides, containing puADA (Figure 3-8). The experiments not only illustrate that puADA can be identified in the context of a typical oligonucleotide but equally important, the control experiment #2 (piperidine treatment of oligos without prior formic acid depurination) indicated no significant autocatalytic depurination of puADA in oligonucleotides.

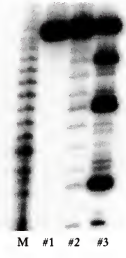


Figure 3-8. Maxam-Gilbert DNA sequencing of oligonucleotides P3-5 ($^{5'}$-ACAGXACTXGGXTT-$^{3'}$), containing puADA. M is marker. The locations of the three nonstandard nucleotides are clearly recognizable. Lane #1 is the starting material, lane #2 represents the piperidine-treated oligonucleotide without prior formic acid treatment. Lane #3 was briefly incubated with formic acid, followed by strand cleavage with piperidine. The differences in migration distance for the sample oligonucleotides in comparison to the oligo T_{4-20} size marker on the left are probably caused by the additional negative charge from puADA. When using lane #2 as an internal marker standard, the fragmentation pattern matches perfectly to the sequence.

Conclusions

A closer inspection of the four lead HIV reverse transcriptase mutants (Y181I, M184V, Y188L, and AZT-21) by primer extension experiments indicated the specific incorporation of

puADA and pyDAD by Y181I and Y188L. Although the polymerase requires up to 24 hours to extend a 50-basepair oligodeoxyribonucleotide with a single nonstandard base in the center region, the standing start experiments and primer extension over two rounds indicate specific incorporation. Furthermore, the efficient primer extension in the standing start experiment and the accumulation of extended primer after incorporation of a nonstandard nucleotide both suggest that the nonstandard bases are incorporated. However, the unusual structural properties such as the C-nucleosidic bond of pyDAD or the hydrogen-donors in the minor groove cause the polymerase to pause (see Chapter 4 – Introduction).

PCR-like amplification of a template containing a single pyDAD was attempted. Preliminary experiments with only standard nucleotides and HIV-RT Y188L lead to successful amplification. However, when switching to the template with nonstandard bases, the extended reaction time caused serious primer degradation. The extended incubation period of 5 days for only 5 cycles not only pushes substrate stability but also allows even small amounts of contaminants to cause significant damage. Such apparent DNA exonuclease activity has not been reported or observed for HIV-RT previously, and we assume that the observed activity has been carried over from the *E. coli* expression system. Experiments by Jeong Ho Park did not succeed in detecting or removing the unwanted contamination (personal communication). Attempts to concentrate the HIV-RT, so that the shorter reaction times may reduce the amount of primer, degraded by the exonuclease failed too. Concentration of the enzyme preparation by membrane filtration increased the enzymatic activity, but also the exonuclease activity. These findings indicate that the rapid purification system applied to our HIV-RT samples is insufficient. Further purification would be required to obtain material suitable for PCR amplification of the pyDAD template.

In summary, the Y188L mutant of HIV-RT was found suitable for specific incorporation of d(puADA)TP opposite pyDAD over the entire pH spectrum (6.5-8.0). Incorporation of d(pyDAD)TP vis-a-vis puADA with acceptable specificity is only possible at pH >7.5. This is largely attributed to the relative high pK_A of pyDAD and the resulting protonation at lower pH, basically depleting the reaction solution of the pyDAD triphosphate available as substrate to the polymerase. When pyDAD serves as template base, no such pH dependency was observed, suggesting that the pH inside the active site of the enzyme may be solution-independent.

A mechanistic explanation for the changes in substrate specificity of Y188L versus wild-type and other mutants of HIV-RT is not known. Although the effects and mechanisms of HIV-RT resistance towards NNRTIs is relatively well understood, relatively little attention has been paid to the implications that such mutations have on the processivity and fidelity of the polymerization step. Steady-state studies on wild-type enzyme in comparison with Y181I and Y188L indicate no significant changes in binding affinity for dNTPs and primer-template complex and the overall reaction rates for the mutants (Debyser et al., 1993). The crystal structure of HIV-RT mutant Y181C (Das et al., 1996) indicates no significant structural changes upon mutation of position 181. A possible interpretation for our results may be the increased flexibility of the polymerase's thumb domain. As mentioned earlier, thumb movement is involved in properly aligning the substrates in the active site, as well as in primer-template translocation. A mutation is the linker region between thumb and palm domain is therefore likely to influence the catalytic cycle of polymerization. Finally, we can not rule out the possibility that most HIV-RT mutants can recognize and incorporate nonstandard bases, sole on their flexibility towards unnatural nucleotides, if the conditions are right.

Even assuming that a fresh batch of Y188L, purified to homogeneity, will successfully amplify the test template with the single nonstandard base, the polymerase's performance is still far from acceptable. Although the elongation time of 24 hours may be reduced to a few hours by using polymerase with a higher specific activity, the faithful and efficient incorporation of multiple nonstandard bases in the next generation templates represents an entirely new challenge. In addition, the fact that HIV-RT is not thermostable may be of secondary importance at this point but needs to be reconsidered eventually.

Furthermore, the difficulties and efforts necessary to come up with Y188L as a polymerase for the very simple task of incorporating one nonstandard base pair indicate the fundamental lack of understanding of the complexity of polymerases. To achieve higher efficiency and improved physiochemical properties for the next generation polymerases, it would be as likely to improve the polymerase's performance by doing the random walk as the attempt of rational protein design. This would in turn require the development of a more efficient system to express, isolate, and assay protein activity. In a first step, the possibilities to apply new purification techniques such as affinity tags to the simultanious isolation of multiple HIV-RT mutants are tested. A novel in vitro assay, based on primer extension and scintillation proximity assay (SPA) is studied in more detail.

CHAPTER 4
PCR WITH PSEUDOTHYMIDINE: HIGH CAPACITY SCREENING OF DNA POLYMERASES USING THE SCINTILLATION PROXIMITY ASSAY

Introduction

In the past, the search for an enzyme that can handle the expanded genetic alphabet has focused on the screening of available wild-type polymerases, as well as their naturally occurring and lab-made mutants. The approach, although partially successful as seen with HIV type-1 reverse transcriptases, has proven extremely slow and technically difficult. In addition, further optimization of the identified mutants is problematic, due to the lack of basic understanding of the structure-function relationship in polymerases. A crude statistical estimate, based on the approximately 200 wild-type and mutant polymerases and reverse transcriptases tested so far (Angerer & Ankenbauer, 1995; Horlacher et al., 1995; Lutz et al., 1998) suggests that an enzyme that performs well with nonstandard nucleotides, can be found. However the search may involves screening of a few thousand polymerase mutants, a number far beyond the technical possibilities of PAGE analysis or the traditional filter-binding assay. A new technology, suitable for screening entire libraries of DNA polymerases, is needed.

In addition to more effective screening techniques, a more detailed structural analysis of the pyDAD/puADA base pair is necessary. To gain a better understanding of why the enzymes perform so poorly with the extra base pairs of the expanded genetic alphabet, a study on the effects of the various modifications that are introduced with the new building blocks

was initiated. Structurally, the pyDAD-puADA base pair has two major differences to a standard base pair: the C-nucleoside character of pyDAD and the hydrogen donors in the minor groove of pyDAD and puADA as shown in Figure 4-1.

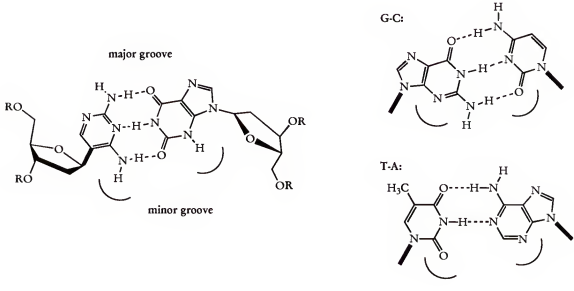


Figure 4-1. Comparison between pyDAD/puADA and the standard base pairs. On the left, the structural distortion through the carbon-carbon bond in pyDAD, as well as the H-donors in the minor groove of either nucleotide is shown. In contrast, both standard base pairs place H-acceptors in the minor groove opposite protein residues (indicated by arcs) for minor-groove scanning.

While the carbon-carbon linkage causes a structural change in the sugar pucker which could result in unusual tertiary structures in an oligonucleotide, the NH groups in the minor groove of both nucleotides could clash with protein side chains from the polymerase that protrude into the minor groove – a mechanism contributing to the overall fidelity of the enzyme and commonly known as "minor groove scanning" (Pelletier et al., 1994; Steitz, 1987).

Structural Effects of C-Nucleosides on Polymerase Performance

As part of this thesis, the effect of C-nucleosides on polymerase performance and fidelity was investigated. Pseudothymidine (ψT, **17**) was used as a model compound for these studies. First isolated from culture filtrates of *Streptomyces platensis* var. *clarensis* (Argoudelis & Mizsak, 1976), ψT is the C-nucleoside analog of thymidine (Figure 4-2). It is synthetically accessible from pseudouridine (see Materials and Methods) and has the advantage of being significantly more stable towards epimerization than pseudouridine, due to the substitution of the NH at 5-position with a methyl group.

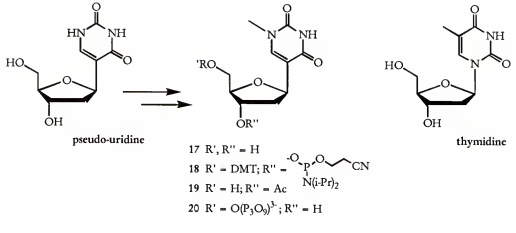


Figure 4-2. Synthesis and structural comparison of pseudothymidine (**17**) and its triphosphate (**20**) with thymidine.

C-Nucleosides vs. N-Nucleosides

While N-nucleosides are the regular DNA and RNA building blocks, C-nucleosides are found naturally only in ribosomal and transfer RNA. Physiologically, C-nucleosides are not incorporated by DNA or RNA polymerases. Instead they are introduced by

posttranscriptional modifications of N-nucleosides (Limbach et al., 1994; and Chapter 1). It is therefore unknown whether C-nucleosides can function as substrates for polymerases. A carbon linkage between the sugar and nucleoside base frequently forces the sugar from a typical 2'-endo into a 3'-endo conformation, pushing the base into a more equatorial position (O'Leary & Kishi, 1994). Once incorporated into DNA, the abnormal sugar pucker conformation may cause the formation of unusual secondary and tertiary structures, which could cause pausing and early termination of the elongation process. Furthermore, while N-nucleosides prefer an anti-orientation, C-nucleosides favor the syn-conformation (O'Leary & Kishi, 1994). The 180° rotation of the base around the glycosidic bond could have an impact on base pairing properties of the unnatural nucleotide. NMR studies on oligoribonucleotides containing pseudouridine indicated that the disturbed sugar pucker (Ψ-effect) is propagated throughout the entire sequence, stabilizing the A-form of the helix (Davis, 1995). On the other hand, NOESY measurements on the same sequence showed that pseudouridine had the proper anti-orientation. Finally, in a previous primer extension study with templates containing ψT and ribonucleotides, Piccirilli et al. (Piccirilli et al., 1991b) reported successful transcription of the C-nucleoside in the presence of T7 RNA polymerases. Unfortunately, no attempts to amplify templates containing C-nucleosides in a PCR-like fashion were described.

In summary, PCR amplification of nucleic acid containing C-nucleosides has never been reported. Pseudothymidine was chosen as our model compound to study the significance of polymerase inhibition by C-nucleosides, one of the possible causes for the failure to incorporate pyDAD by enzymatic reactions.

High-Throughput Screen for DNA Polymerases

The search for polymerases that incorporate pyDAD and puADA requires the screening of thousands of enzymes. Previously employed techniques such as primer extension reactions, followed by PAGE analysis or the filter binding assay are too complicated to allow efficient screening of large sample numbers.

Recognizing the limits of these methods, Loeb and coworkers recently reported the development and successful testing of an in vivo assay for DNA polymerases (Kim & Loeb, 1995; Suzuki et al., 1996; Sweasy & Loeb, 1992). The approach takes advantage of a mutation in the *E. coli* polymerase I gene that makes the host enzyme thermosensitive (Witkin & Roegner-Maniscalco, 1992). Upon raising the growth temperature from 30°C to 37°C, only strains that carry a functional DNA polymerase in a plasmid will be able to continue to proliferate and survive. A library of randomly mutated polymerases can therefore be easily screened by simply transforming a plasmid library of the target polymerase into this *E. coli* strain and growing the colonies at 37°C. Only strains with a complementary polymerase will survive and thus be isolatable from the agar plates.

Despite its elegance, the method has several limitations. The growth temperature of 37°C for *E.coli* limits the testing to enzymes from mesophilic organisms. *Taq* DNA polymerase was the only thermostable enzyme successfully tested with the assay (Suzuki et al., 1996). However, the method was applicable only because *Taq* shows some activity at 37°C, a property unique to *Taq* among all known thermostable DNA polymerases. Experiments with the assay have also indicated a relatively high rate of reversion of the *E. coli* DNA polymerase, the loss of thermosensitivity of the host enzyme. This raises significantly the background level for the assay (J.H. Park – personal communication). Most important, however, the technique can not be used to select directly for incorporation of pyDAD and puADA. Such in vivo

experiments would require the cellular uptake of nonstandard nucleosides from the media. Following the membrane transport, the building blocks would then have to be phosphorylated by kinases to the triphosphate. Studies on phosphorylation of nonstandard nucleotides with Herpes Simplex Virus (HSV-1) thymidine kinase, known for its broad substrate specificity, in collaboration with Dr. G. Folkers laboratory at the University of Zurich showed very efficient activation of d(puADA). In contrast, however, activation of d(pyDAD) failed completely (B. Pilger, G.Folkers, S.Lutz, S.A.Benner - unpublished results). Another concern for in vivo studies is that puADA (xanthosine) is part of the synthetic pathway for guanosine. Large concentrations of puADA may affect the cell's regulatory mechanism, resulting in rapid degradation of the nucleoside inside the cell.

Another, recently published, intracellular approach using phage display to identify functional polymerases is not applicable for similar reasons (Jestin et al., 1999). In summary, in vivo techniques to screen for functional polymerases and reverse transcriptases are available but at this point not applicable to search for enzymes that can utilize the expanded genetic alphabet. A new in vitro methodology therefore needed to be developed.

Scintillation Proximity Assay

The scintillation proximity assay (SPA™, Amersham/Pharmacia Biotech, Arlington Height, IL) was recently introduced as an alternative to polyacrylamide gel electrophoresis and filter-binding assays for measuring polymerase activity (Cole, 1996). In the regular primer extension experiment, tritium-labelled dNTP is added to the reaction mixture together with 5'-biotinylated primer. Following an elongation period, streptavidin-coated plastic beads containing scintillant are added to the reaction. The 5'-biotinylated primer/ template complex binds with high affinity, placing the incorporated tritium-labelled nucleotides in close

proximity to the plastic bead. Because of the relatively low energy of tritium decay particles, only isotopic sources in the vicinity of the plastic bead result in a detectable signal, which therefore represents the amount of labelled dNMP incorporated. Unincorporated [^{3}H]-dNTP, still present in solution, accounts only for background counts.

An analogous strategy can be implemented using Flashplates (NEN™ Life Science Products, Boston, MA), a microtiter plate where the streptavidin is on the surface of the plastic wells and scintillant is embedded in the plastic. In summary, minimal workup of a primer extension experiment is required when used with the SPA. The assay allows the in vitro screening of up to 96 polymerases at once.

Materials and Methods

Experimental Procedure - Chemical Synthesis

Reagents and solvents for chemical synthesis were purchased from Aldrich Chemicals and were of the highest available quality. Solvents were dried by distillation in the presence of the appropriate drying agents (Perrin & Armarego, 1988). For details on the instrumentation see Chapter 2 (Materials and Methods).

5'-[O-(4,4'-Dimethoxytrityl)-2'-deoxy-ψ-thymidine]-3'-[2-cyanoethyl-bis(isopropyl)] phosphoramidite (18). Compound **18** was prepared from pseudouridine via pseudothymidine (17) (Bhattacharya *et al.*, 1995). The final product was obtained in nine steps with an overall yield of 25% and characterized extensively. ^{1}H-NMR (300 MHz, $CDCl_3$) δ: 8.77 (s, br, 1H, NH), 7.46-7.40 (m, 3H, H-6, arom.), 7.34-7.20 (m, 7H, arom.), 6.88-6.80 (m, 4H, arom.), 5.09 (t, 1H, H-1', J= 7.2 Hz), 4.52 (m, 1H, H-3'), 4.11 (m, 1H, H-4'), 3.86-3.70 (m, 1H, H-

(CH_2CN)), 3.79 (d, 6H, H-(OCH_3)), 3.67-3.48 (m, 3H, H-5'$_{A/B}$, H-(CH_2CN)), 3.36 (m, 1H, H-(CH_2CN)), 3.22 (m, 1H, H-(CH_2CN)), 3.08 (d, 3H, H-(NCH_3)), 2.62 (t, 1H, iPr), 2.55 (m, 1H, H-2'), 2.44 (t, 1H, iPr), 1.97 (m, 1H, H-2'), 1.28-1.23 (m, 12H, iPr), ppm. ^{13}C-NMR (300 MHz, $CDCl_3$) δ: 162.4 (C-2), 158.6, 150.8 (C-4), 144.7, 141.2 (C-6), 135.8, 130.1, 128.2, 127.8, 127.0, 115.7 (d, C-5), 113.1, 103.3, 86.2 (C-1'), 84.8 (C-4'), 73.4, 63.4 (d), 58.2 (d), 55.2, 43.2 (d), 40.8 (d), 35.7, 24.5 (iPr), 20.2 (iPr), ppm. NOE (300 MHz, $CDCl_3$) δ: IRR at 5.09 (H-1'): 7.42 / 7.45 / 4.11 (H-4') / 2.55 (H-2'), ppm. ^{31}P-NMR (300 MHz, $CDCl_3$) δ: 148.5, 147.8, ppm (against phosphoric acid in deuterated water as external standard). LCMS FAB (m-NPA): 745 ($M^+ + 1$). TLC: dichloromethane/ethyl acetate/triethyl-amine 50:50:1, R_f = 0.43, 0.34.

3'-Acetyl-pseudothymidine (19). Compound **17** (250 mg; 1.03 mmol) was coevaporated twice with pyridine and dissolved in pyridine (5 mL). Separately dimethoxytrityl chloride (420 mg; 1.24 mmol; 1.2 eq.) was dissolved in pyridine (5 mL). The dimethoxytrityl chloride solution and triethylamine (0.32 mL) were added simultaneously to the solution of **17** and the mixture shaken at room temperature overnight. The reaction was quenched by addition of methanol (0.3 mL) and the solvent evaporated. The residue was coevaporated twice with toluene and the dried residue was dissolved in chloroform, followed by two washes with saturated sodium bicarbonate solution. The combined organic layers were dried over sodium sulfate and the solvent evaporated. The yellow foam was purified by silica gel column chromatography (chloroform/methanol 15:1), yielding 470 mg (84%) of the 5'-protected nucleoside.

To a solution of 5'-protected nucleoside (170 mg; 0.31 mmol) in pyridine (5 mL) was dropwise added acetic anhydride (120 μL; 1.25 mmol; 4 eq.). The solution was shaken at room temperature for 4 h. The reaction was quenched with methanol (0.6 mL). The solvent was evaporated and the raw material dried under vacuum overnight. The white-yellow foam was

dissolved in dichloromethane (15 mL) and trifluoroacetic acid (0.35 mL) was added dropwise. The solution turned dark red and was shaken at room temperature for 20 min. After quenching with saturated sodium bicarbonate (10 mL), the aqueous layer was extracted extensively with chloroform and the combined organic layer dried over sodium sulfate. The filtrate was evaporated and the residue purified over silica gel (chloroform/methanol 10:1), yielding 33 mg (38%) of **19**. ^{1}H-NMR (300 MHz, MeOH-*d*) δ: 7.69 (s, 1H, H-6), 5.20 (d, 1H, H-3', $J_{H3/H4}$ = 6 Hz), 4.89 (dd, 1H, H-1', $J_{1/2}$ = 5.1, 10.5 Hz), 4.00 (q, 1H, H-4', $J_{H4/H5}$ = 5.5 Hz, $J_{H4/H5}$ = 3.3 Hz), 3.72 (ddd, 2H, H-5'$_{AB}$, J = 3.3, 4.8 Hz), 3.35 (s, 3H, N-CH$_3$), 2.30 (dd, 1H, H-2'$_B$, $J_{H2/H1}$ = 5.1 Hz, $J_{H2/H2}$ = 10.5 Hz), 2.07 (s, 3H, Ac), 2.06 (m, 1H, H-2'$_A$), ppm. ^{13}C-NMR (300 MHz, MeOH-*d*) δ: 172.4, 165.3, 144.8, 114.0, 86.7, 78.2, 76.1, 63.6, 39.7, 36.1, 21.0, ppm. TLC: chloroform/methanol 10:1, R_f = 0.32.

Pseudothymidine-5'-triphosphate (20). Compound **19** (8.5 mg, 30 μmol) was dissolved in a mixture of pyridine and dioxane (30 μL / 89 μL) and reacted with 2-chloro-4H-1,3,2-benzo-dioxaphosphorin-4-one (36 μL; 188 mg in 0.93 mL dioxane). After 10 min, tributylammonium pyrophosphate (17 mg; 37 μmol) in DMF (91 μL) and tributylamine (31 μL) were added and the reaction mixture shaken for another 10 min. The reaction was quenched by addition of iodine solution (0.6 mL; 1% in pyridine/water 98:2). After shaking the mixture at room temperature for 15 minutes, the excess iodine was destroyed with a few drops of Na_2SO_3 solution (5%). The solution was evaporated and dried under vacuum. The residue was hydrolyzed with water (0.8 mL) at room temperature for 30 minutes. The solution was lyophilized and the residue incubated in concentrated ammonia (1 mL) at 55°C overnight. After degassing and lyophilization, the crude product was purified by RP-HPLC (Waters Prep. Nova Pak HR C18, 7.8 x 300 mm; 4 mL/min; 10 mM TEAA (pH 7.0) - 10 mM TEAA (pH 7.0) + 20% acetonitrile; 0 to 25 % B in 30 minutes). The product fractions were pooled and

lyophilized three times to remove excess TEAA, yielding **20** (2 mg, 14%) as a white-yellow foam. ^{1}H-NMR (300 MHz, D_2O) δ: 7.85 (s, 1H, H-6), 5.18 (q, 1H, H-1', J = 6.04, 9.61 Hz), 4.62 (hept, 1H, H-3', J = 2.5, 5.3 Hz), 4.20 (m, 1H, H-4'), 3.71 (dd, 1H, H-5'$_A$, J = 4.4, 11.7 Hz), 3.62 (dd, 1H, H-5'$_B$, J = 6.3, 11.7 Hz), 3.47 (s, 3H, NCH_3), 3.27 (q, triethylammonium), 2.34 (ddd, 1H, H-2'$_A$, J = 2.47, 6.05, 13.5 Hz), 2.18 (ddd, 1H, H-2'$_B$, J = 5.7, 9.61, 13.5 Hz), 1.35 (t, triethylammonium), ppm. ^{31}P-NMR (300 MHz, D_2O) δ: 3.05 (d, γP, J = 18.5 Hz), 2.23 (d, αP, J = 24.4 Hz), -9.73 (t, βP, J = 18.5, 24.4 Hz), ppm. LRMS ESI: 503 (M^- +Na), 481 (M^- - 1). UV (Tris-HCl, 100 mM, pH 7.0): λ_{max} = 271 nm (ε = 8800 $M^{-1}cm^{-1}$).

DNA synthesis and digestion. Oligodeoxyribonucleotides containing **17** were prepared on an Applied Biosystems DNA synthesizer. The coupling times for **18** were doubled. No further adjustments of the synthesis protocol were necessary. The oligonucleotides were deprotected under standard conditions (ammonia solution, 55°C, overnight) and PAGE-purified (see Appendix C8 and C10).

Successful incorporation of **17** by chemical synthesis or enzymatic primer extension was shown through enzymatic digestion of the oligonucleotides with phosphodiesterase (*Crotalus durissus terrificus*) and alkaline phosphatase (*Bovine calf intestine*). For details see appendix C11. Subsequent RP-HPLC analysis (5% acetonitrile in TEAA (10 mM, pH 7) over 15 min) was used to quantify the composition of the oligonucleotides (**17**; λ_{max} = 271 nm, ε = 6000 $M^{-1}cm^{-1}$, $t_R(\psi T)$ = 10-11 min).

Primer and templates. The following sequences were used for the various primer extension experiments and PCR amplifications. Template T4-4 was synthesized to demonstrate the successful incorporation of multiple **17** (X) in an oligonucleotide. Primer P4-2 was also prepared with a 5'-biotinylated substitution (P4-2b).

P4-1: 5'-GCG AAT TAA CCC TCA CTA AAG-3' (T3 primer sequence)

P4-2(b): 5'-GCG TAA TAC GAC TCA CTA TAG-3' (T7 primer sequence)

P4-3: 5'-GCG TAA TAC GAC TCA CTA TAG CGC-3'

T4-1: 5'-GAG ACX GCG CTA TAG TGA GTC GTA GGA CGC-3'

T4-2: 5'-GAG ACA GCG CTA TAG TGA GTC GTA GGA CGC-3'

T4-3: 5'-GCGAATTAACCCTCACTAAAGTACGTTCGTCTATAGTGAGTCGTATTACGC-3'

T4-4: 5'-GCGAAXXAACCCXCACXAAAGXACGXXCGXCTATAGTGAGTCGTATTACGC-3'

Experimental Procedure - Molecular Biology

Polymerase screening by primer extension and scintillation counting. A series of commercially available thermostable polymerases was tested with the primer extension / SPA assay: *Taq* DNA polymerase and *Tth* DNA polymerase (Promega), *Pfu* DNA polymerase (cloned; Stratagene), *Pwo* DNA polymerase (Boehringer Mannheim), Vent™, Vent (exo-)™, Deep Vent™, Deep Vent (exo-)™, and 9°N DNA polymerase (all from New England Biolabs). Each enzyme was incubated in 5 separate reactions (Figure 4-3) for 30 minutes at 72°C with a mixture containing the supplier's reaction buffer, 1 μM of each dNTP (Promega) as well as either [^{3}H] TTP (Amersham, 90-130 Ci/mmol) or [^{3}H] dATP (ICN, 23 Ci/mmol), and 0.2 pmol of the primer/template complex (P4-2b/T-1) (total volume: 50 μL). The reaction was quenched with EDTA (20 μL, 0.5 M, pH 8.0), and the mixture was mixed with SPA™-beads (Amersham) or loaded into a Flashplate™ (NEN). After incubation at room temperature for 15-30 minutes, the samples were counted in a Topcount NXT Scintillation Microtiterplate Reader (Packard Instruments, Meriden, CT).

Standing start experiments. The qualitative fidelity of *Taq* and *Tth* DNA polymerase was investigated by standing start experiments. Primers with a 5'-^{32}P-label were obtained by enzymatic phosphorylation with polynucleotide kinase (Appendix C7) and were ethanol-precipitated twice. After mixing primer P4-3 (0.5 μM), template T4-1 or T4-2 (0.625 μM), the corresponding dNTPs (200 μM), and reaction buffer (supplied by manufacturer), the enzyme (2.5 U per 8 μL reaction mixture) was added on ice and the sample cycled once (1 min at 94°C; 30 sec at 55°C; 10 min at 72°C). The reaction was quenched with EDTA (0.5 M, pH 8) and ethanol-precipitated before loaded on a 10% PAGE gel (7 M urea). The gel was analyzed with the MolecularImager® (BioRad, Hercules, CA).

PCR amplification. For PCR amplification, primer P4-1 and P4-2 (1 μM each, P4-2 was 5'-^{32}P labelled) and template (T4-3; 20 nM), as well as the appropriate dNTPs (200 μM each), were mixed with reaction buffer and *Taq* DNA polymerase (25 U) and adjusted to a final volume of 100 μL with water. The experiments were cycled (initial: 5 min, 94°C// 30 sec., 94°C; 30 sec., 55°C; 10 min, 72°C// polishing: 5 min, 72°C) five to ten times. During each annealing phase, an aliquot (8 μL) was taken, quenched with EDTA (0.5 M, pH 8), and precipitated with ethanol. The samples were separated on a 10% PAGE gel (7 M urea) and analyzed with the MolecularImager® (BioRad, Hercules, CA).

DNA sequencing. The accuracy of the PCR amplification was analyzed by Sanger dideoxy-sequencing with *Taq* DNA polymerase (Innis et al., 1988) and chemical sequencing using the protocol of Maxam & Gilbert (Sambrook et al., 1989). Single-stranded oligonucleotides for DNA sequencing were produced by substituting P4-2 with the 5'-biotinylated P4-2b. After 10 cycles of PCR (same protocol as above), the reaction was quenched with EDTA (0.5 M, pH 8) and loaded onto a preequilibrated streptavidin-agarose column (Fluka Chemicals), washed with 6 volumes of Tris-HCl (10 mM, pH 7.5), NaCl (50

mM) , EDTA (1 mM, pH 8), and the unbiotinylated strand was eluted with 2 volumes 0.2 M NaOH. After ethanol-precipitation and PAGE purification, the single-stranded material was used for Sanger dideoxy and Maxam-Gilbert sequencing.

Sequence analysis by the dideoxy method required some adjustment of the experimental conditions. When the temperature of the primer extension reaction was set at 70°C, the following dNTP/ddNTP/Mg concentrations were found suitable to sequence short oligonucleotides with *Taq* DNA polymerase: dA/ddA/Mg 15 μM / 2 mM / 2.24 mM; dT/ddT/Mg 15 μM / 3 mM / 3.24 mM; dG/ddG/Mg 15 μM / 0.5 mM / 0.74 mM; dC/ddC/Mg 15 μM / 1 mM / 1.24 mM. After annealing equimolar amounts of ^{32}P-labelled primer (P4-2) with the purified PCR product, primer/template complex (0.25 μM) was incubated with the dNTP/ddNTP/Mg mixtures and 10 U of *Taq* DNA polymerase for 10 minutes.

For the chemical sequence analysis, the protocol by Maxam and Gilbert for cleavage after purines and pyrimidines was employed. After 5'-^{32}P labelling of the amplified oligonucleotide with T4 polynucleotide kinase (Appendix C7), the material was ethanol-precipitated twice to remove all traces of triphosphates. Approximately 100,000 cpm (2 μL) of the labelled PCR product were used per derivatization. Purine bases (A + G) were depurinated in formic acid (2.3%) for 20-30 min at 37°C, followed by strand cleavage with piperidine (150 μL, 1 M, 90°C, 30 min). Pyrimidine bases were derivatized by hydrazine treatment (23 μL, 20°C, 4-10 min), quenched with hydrazine stop solution (200 μL) (Sambrook et al., 1989) and ethanol (750 μL). Following lyophilization, the derivatized pyrimidine bases were cleaved with piperidine (100 μL, 1 M, 90°C, 30 min). All reaction mixtures from the dideoxy sequencing, as well as from the Maxam-Gilbert cleavage, were electrophoretically resolved by PAGE (10%, 7 M urea) and analyzed by autoradiology in the MolecularImager®.

Gel fidelity assay for ψT. Once a polymerase that accepted 17 was identified by the SPA, kinetic data for the incorporation of 17 by the polymerase were determined by single nucleotide incorporation (Boosalis et al., 1987). The same primer/template complexes as in the previous qualitative analysis (standing start experiments) were used. In the absence of dNTP, primer P4-3 (^{32}P-5'-labelled; 20 pmol), template T4-1 or T4-2 (25 pmol), and *Taq* DNA polymerase (0.1 U) were mixed in reaction buffer (final volume: 8 μL), denatured at 94°C (1 min), and annealed at 55°C (60 min). The experiment was initiated upon addition of the appropriate dNTP (2 μL) at 55°C and the reaction was quenched after the previously determined time to ensure that a maximum of 20% of the primer was extended (dA-template: TTP = 180 s, ψTTP = 240 s, dATP = 20 min; ψT-template: dATP = 300 s, dGTP = 20 min). For each dNTP, the concentration in the reaction mixture was varied from 3.12 μM to 200 μM (in 7 steps by 1:1 serial dilution). After separation by PAGE, the products were quantified using the MolecularImager® and MultiAnalyst® software (BioRad). Kinetic parameters were determined by plotting [dNTP] versus [dNTP]/v (v is the primer extension velocity (ratio $(n+1)/(n+(n+1))$ per second, n is the pixel counts per band) (Hanes-Woolf plot). From the linear graph, V_{max} can be calculated from the slope ($1/V_{max}$); K_M was determined from the y-intercept (K_M/V_{max}).

Results and Discussion

In Vitro Screening Assay

The combination of primer extension experiments with SPA was tested to screen for DNA polymerases able to incorporate dATP opposite pseudothymidine (**17**, ψT), the C-nucleoside analog of thymidine, in a template.

Synthesis of pseudothymidine

Pseudothymidine **17** was prepared from the commercially available pseudouridine and converted to its phosphoramidite as described in the experimental section. Oligonucleotides containing **17** were prepared by phosphoramidite-based DNA synthesis. The successful integration of ψT was proven by enzymatic digestion of the oligonucleotides. In addition, the triphosphate **20** was prepared for enzymatic studies with DNA polymerases.

Primer extension experiments

The primer extension experiments are illustrated schematically in Figure 4-3. The template (T4-1) was designed to contain no 2'-deoxyadenosine between the 3'-end of the primer (position 21) and **17** (X; position 25). After annealing with a 5'-biotinylated primer (P4-2b), the complex was incubated with the polymerase and the standard dNTPs in the presence of [^{3}H]-TTP. Successful incorporation of dATP opposite **17** permits further elongation of the primer beyond position 25, and consequently incorporation of tritium-labelled TTP at position 27. The successful incorporation of the radioisotope could then be detected by the SPA. A series of control experiments would ensure that the observed effect was actually the result of primer extension.

To learn whether the polymerase extended the primer past **17** in the template via misincorporation, an analogous experiment was run in the absence of dATP (Figure 4-3, reaction #2). While experiment #1 tested for primer extension beyond the unnatural nucleotide, experiment #2 measured the specificity of the incorporation in the absence of dATP (the appropriate hydrogen bonding partner for **17**). Meanwhile, experiment #3 detected the background radioactivity of the reaction mixture. The incorporation of the correct

nucleoside triphosphate opposite 17 can also be measured by using tritium-labelled dATP in a second, independent set of reactions (Figure 4-3, reaction #4/5).

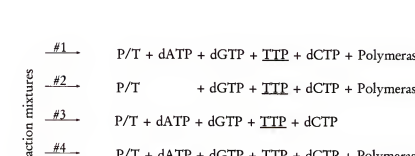


Figure 4-3. Summary of primer extension experiments. Each polymerase was tested on the primer/template complex (P4-2b/T4-1) in five reaction mixtures. The underlined dNTP is tritium-labelled and functions as a reporter for successful incorporation.

A set of commercially available thermostable polymerases suitable for the polymerase chain reaction was screened for their ability to incorporate 17 using the P4-2b/T4-1 complex. Five reactions were run for each polymerase. After quenching with EDTA, the amount of tritium incorporation was determined by scintillation counting on SPA beads (Figure 4-4). In separate experiments, the incorporation was shown to be linear as a function of time and the concentration of polymerase, with results reproducible to within 10%.

Results from the primer extension experiments. From the data in Figure 4-4, several questions concerning polymerase performance in the presence of unnatural nucleotides can be addressed. First, the correlation between the data from reaction #1-3 and #4-5 (Figure 4-3, 4-4) must be discussed. Given the specific activity of [^{3}H]-TTP (~118 Ci/mmol) and [^{3}H]-dATP

(~ 16 Ci/mmol) used in the experiments, a polymerase that successfully completes elongation with no mismatching should generate ca. 7 times more radioactivity in the products in experiments #1-3 than in experiments #4-5. A lower ratio indicates failure to complete incorporation past the initial dA to yield full-length product. In Figure 4-4, the amounts of label introduced in the parallel experiments with [^{3}H]-TTP and [^{3}H]-dATP approached the ratio of 7:1 for *Taq* and Deep Vent DNA polymerases only. The ratio was lower for the other polymerases examined, suggesting that they stalled and aborted elongation after [^{3}H]-dATP incorporation.

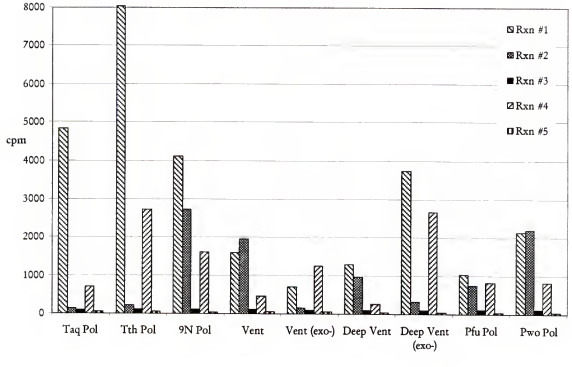


Figure 4-4. Graphical presentation of primer extension by various thermostable DNA polymerases. For each enzyme, all five experiments (#1 - #5; Figure 4-3) were performed and the amount of tritium incorporation was measured by liquid scintillation counting (SPA).

Second, the data in Figure 4-4 show a surprisingly high, direct correlation between the amount of products containing [^{3}H]-TMP synthesized in the absence of dATP (experiment #2) and the exonuclease activity of the polymerase. Thus, Vent, Deep Vent, Pfu, and Pwo DNA polymerases, all having high exonuclease activity, all displayed a high ratio of signal in experiment #2 relative to that in experiment #1. This could, of course, arise by primer extension involving misincorporation opposite **17** or skipping over **17**, followed by full-length extension. Another possibility could be the incorporation of [^{3}H]-TMP opposite **17**. More likely, however, is the hypothesis that an exo$^+$ polymerase stalls at the unnatural nucleotide in the template. At this point, the exonuclease activity digests the primer backward, eventually exposing the adenosine at position 19 in the template. Upon switching back into polymerization mode, the polymerase then directs the incorporation of [^{3}H]-TMP opposite the adenosine in position 19. The digestion of the primer is well known with exo$^+$ polymerases (Derbyshire et al., 1995), especially when triphosphate concentrations are as low as those used here.

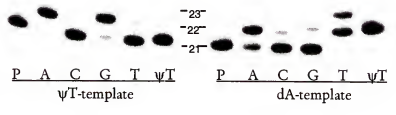
Based on the screening using the scintillation proximity assay, the DNA polymerases from *Thermus aquaticus* and *Thermus thermophilus* were chosen for standing start experiments (Boosalis et al., 1987). *Taq* DNA polymerase because of the high ratio of [^{3}H]-TMP to [^{3}H]-dAMP incorporation in the parallel experiments #1 and #4, *Tth* DNA polymerase because of an apparently high fidelity.

Standing start experiments

For the standing start experiments, two primer/ template complexes (P4-3/T4-1 and P4-3/T4-2) carrying either **17** or 2'-deoxyadenosine at position 25 were prepared. An extended reaction time of 10 min allowed the polymerase to undergo several cycles of polymerization

(multiple hit conditions). The resulting gel (Figure 4-5), while not suitable for quantitative analysis, identified nucleoside triphosphates that might be misincorporated opposite 17 in the template, as well as the competitors of ψTTP (20) for incorporation opposite dA in the template.

A. *Taq* DNA polymerase experiments:



B. *Tth* DNA polymerase experiments:

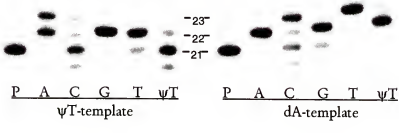


Figure 4-5. Standing start experiments with *Taq* (**A**) and *Tth* (**B**) DNA polymerase. Radiolabelled primer (P4-2, 21-mer) was annealed to T4-1 and incubated with the indicated nucleoside triphosphate and 2.5 U of DNA polymerase. Each reaction was run for 10 minutes at 72°C. In addition, P4-2 was run as a control (P). For further experimental details see Materials and Methods.

Tth DNA polymerase showed significant misincorporation of purines and thymidine opposite 17. In the opposite direction, having dA in the template, the polymerase incorporates all the triphosphates, indicating almost zero specificity. These data, obviously in disagreement with the SPA results, indicate extremely poor performance of *Tth* DNA polymerase under the

given conditions. The differences observed are probably the result of the high dNTP concentrations (200 μM) in the absence of the other dNTPs, conditions unfavorable for high specificity of *Tth* DNA polymerase. No further experiments under optimized conditions were done, considering the excellent performance of *Taq* DNA polymerase (Figure 4-5A). The DNA polymerase from *Thermus aquaticus* extended the primer by placing a purine nucleotide, preferentially 2'-deoxyadenosine, opposite to **17** in the template. *Taq* DNA polymerase also incorporated ψTTP efficiently and specifically opposite dA in the template. The only nucleoside triphosphate misincorporated to a significant extent was dATP.

Steady-state kinetics of ψT-incorporation by gel fidelity assay

A quantitative comparison of the efficiency of incorporation of ψT and its natural competitors was made by measuring the single-nucleotide insertion by varying the dNTP concentration under single-turnover conditions. Applying the same primer-template complexes as above (P4-3/T4-1 and P4-3/T4-2), a series of kinetic studies was carried out using steady-state methods (Creighton et al., 1995; Goodman et al., 1993). Based on the qualitative findings in the standing start experiments, kinetic data were determined only for the possible competitors and **17** itself. The insertion of dATP and dGTP opposite ψT in the template and the incorporation of TTP, ψTTP, and dATP opposite dA were studied. Kinetic parameters derived from the data are presented in table 4-1. The insertion of dATP opposite ψT in the template is 50-fold more efficient than insertion of dGTP. In the opposite direction, incorporation of ψTTP vis-à-vis dA in the template is only twofold less efficient than TTP, the natural hydrogen bonding partner of dA. The insertion of dATP, on the other hand, is 3 $x10^2$-fold lower and probably the result of non-template specific polymerization.

Table 4-1. Kinetic parameters for ψT-incorporation by *Taq* DNA polymerase

Template	dNTP	V_{max}	K_M	Efficiency	Accuracy
		(% min^{-1})	(μM)	(V_{max}/K_M)	
ψT	dATP	3.3 (±0.4)	17.7 (±2.6)	1.8×10^5	1
	dGTP	0.2 (±0.01)	51.3 (±2.1)	3.8×10^3	0.021
dA	TTP	9.7 (±0.2)	9.0 (±0.2)	1.1×10^6	1
	ψTTP	6.1 (±0.1)	13.6 (±1.7)	4.5×10^5	0.41
	dATP	0.7 (±0.01)	183 (±0.2)	3.7×10^3	0.003

Steady-state parameters for a single nucleotide incorporation into (P4-2/T4-1 and P4-2/T4-2). The optimal reaction times were determined in a time-course experiment (see experimental data). To determine the kinetic data, the dNTP concentration was varied from 3.12 μM to 200 μM (a total of 7 concentrations). All experiments were run three times and the averaged data are shown. Error limits are given in parentheses.

PCR amplification

Once performance with good fidelity and efficiency under single nucleotide incorporation conditions was achieved, primer elongation by *Taq* DNA polymerase under PCR-like conditions was tested. A reaction mixture containing a standard oligonucleotide template (T4-3), as well as the two flanking primers (P4-1, P4-2), was prepared and PCR amplified in the presence of dATP, dCTP, dGTP, and ψTTP. Two control reactions, one substituting ψTTP with standard TTP, the other performed in the absence of any thymidine analog, were run simultaneously. A sample was taken after each cycle and analyzed by PAGE (Figure 4-6). The reaction mixture lacking thymidine or its analog generated no full-length product; the primer extension reaction was terminated completely at the position of the first adenosine in the template. In the presence of ψTTP, the desired full-length oligonucleotide represented the primary product. Although up to three ψT's in succession need to be

incorporated to produce full-length product, only a small fraction of the extended primer appeared as an early termination product.

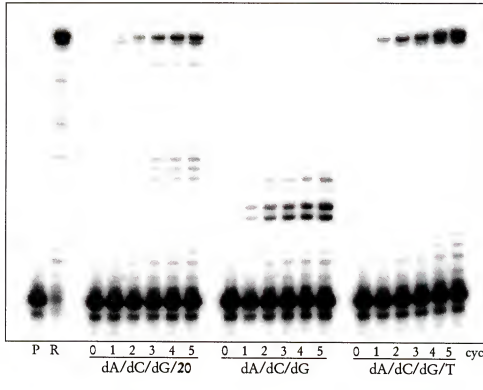


Figure 4-6. PCR-like amplification over 5 cycles. Amplification during each cycle required incorporation of 13 thymidines or thymidine analogs to obtain full-length product. Positive control R (^{32}P-labelled template) and primer P are shown.

In addition, a PCR experiment with **20** over 10 cycles was performed. Ten rounds are the minimum number of cycles required for the amplification step in an in vitro selection experiment (P.Burgstaller – personal communication). Using the standard template T4-3 with primer P4-1 and P4-2, the successful elongation was monitored by PAGE analysis of small aliquots, taken after each round during the annealing phase. The resulting gel in Figure 4-7 presents the successful amplification of full-length product. Small amounts of early termination product can be seen, in particular in the region of three consecutive dA's in the template.

Integration of the overall counts per lane however indicate that more than 90% of the extended primer are full-length product.

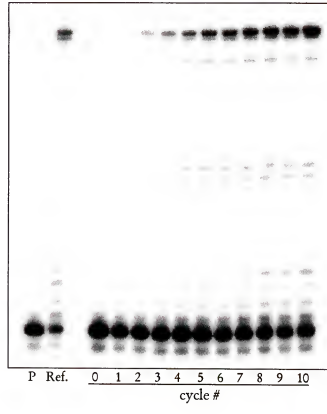


Figure 4-7. Ten rounds of PCR amplification of T4-3 in the presence of P4-1 and 5'-^{32}P labelled P4-2, as well as dATP, dCTP, dGTP and ψTTP (20).

Amplification fidelity and sequencing

In addition to the evidence of ψT-incorporation by *Taq* DNA polymerase under conditions of PCR amplification, further proof for sequence-specific amplification of an oligonucleotide in the presence of ψTTP was sought. Modifying the original PCR protocol, primer P4-2 was replaced by its 5'-biotinylated analog P4-2b. After PCR amplification over 10

cycles, the reaction mixture was loaded onto a streptavidin column. After extensive washing, the bound double-stranded DNA's were denatured with elution buffer (sodium hydroxide solution) and the unbiotinylated single-stranded template recovered from the elution buffer by repeated ethanol precipitation. The excess primer was removed by preparative PAGE and the resulting PCR product was analyzed by enzymatic digestion, as well as enzymatic and chemical sequencing.

A complete digestion of the PCR product to give the overall nucleoside composition by phosphodiesterase and alkaline phosphatase (Appendix C11), followed by quantitation by RP-HPLC, gave the predicted ratio of nucleosides. Aliquots of the PCR product were then used for Sanger dideoxy-sequencing and Maxam-Gilbert sequencing (Figure 4-8A/B).

In the Maxam-Gilbert experiment, the positions of ψT could be identified by the missing bands in the C+T patterns. The C-nucleoside does not undergo strand cleavage upon hydrazine/piperidine treatment and therefore does not appear as a distinct band in the sequencing gel. Further confidence for a faithful replication of A-ψT base pair by *Taq* DNA polymerase arises from the absence of a ψT to purine transition. Based on the single nucleotide incorporation experiments (Figure 4-5), adenosine and guanosine are most likely to be misincorporated opposite ψT or instead of ψTTP. Such a transition would easily be identified by an additional band in the A+G bands pattern. The Sanger dideoxy sequencing of the PCR product confirmed sequence identity with the original template (Figure 4-8B). Excluding the first few positions after the primer, which could not be read conclusively due to high background, the sequence pattern was consistent with the original template.

In summary, both experiments show that the sequence patterns of the PCR-amplified product were indistinguishable from the pattern of the original template. These results show a high level of fidelity at the individual steps in the PCR reaction.

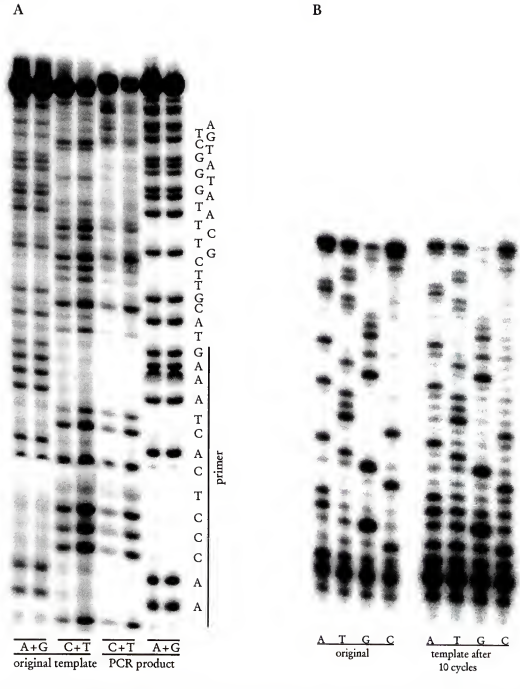


Figure 4-8. Sequencing of PCR-generated oligonucleotides, amplified using ψTTP instead of TTP. A) Chemical sequencing by Maxam-Gilbert method. Each sequence was incubated for two different time periods (A+G: 21 and 30 minutes; C+T: 4 and 10 minutes). The depurination pattern (A+G) for the original template and the PCR product are identical. The cleavage reaction for (C+T) is not as efficient. The sequencing pattern showed no evidence for substantial replacement of ψT by purines. B) Sanger dideoxynucleotide sequencing. The sequencing patterns from the original template and the PCR product are indistinguishable. The additional weak bands are attributed to background termination of *Taq* DNA polymerase and extended exposure time.

Conclusions

We have successfully tested a new in vitro screening assay for thermostable DNA polymerases. The combination of a set of primer extension reactions and the SPA* technology for its rapid and simple quantification represents a novel method to search for polymerases that can utilize functionalized or structurally modified nucleotides. The method is applicable to any nucleoside analog, whether integrated in an oligonucleotide or present as the triphosphate. The methodology can now be extended to screen polymerases for their ability to faithfully deal with nonstandard base pairs. The approach becomes particularly interesting for the search of protein sequence space by directed evolution; an approach that introduces random mutations on the genome level, followed by expression and screening of the resulting protein libraries. The approach will be discussed in the following chapters.

For pseudothymidine, the best polymerase identified by primer extension and SPA was *Taq* DNA polymerase. After the preliminary screening, more detailed kinetic studies were performed to demonstrate that the polymerase did perform as predicted based on the results from the SPA. Although the fidelity of synthesis of the A-ψT base pair by *Taq* DNA polymerase is a factor of two lower than with the conventional base pairs, it is sufficient for the amplification of typical oligonucleotides for in vitro selection (up to 150 nucleotides over 10 cycles). The PCR reaction with pseudothymidine was the first example of PCR-like amplification with a C-nucleoside.

Based on the results from the kinetic studies, the significance of hydrogen bonding during polymerization can be derived. The data from the single nucleotide insertion can be put into perspective with previously published data on related nucleoside analogs such as difluorotoluene deoxynucleoside (dF, (Moran et al., 1997a; Moran et al., 1997b)). Kool and

coworker report therein the successful enzymatic incorporation of a C-nucleoside analog that in addition lacks hydrogen bonding capacity. Although the controversy around the ability of fluorine to form hydrogen bonds is still ongoing (Diederichsen, 1998; Evans & Seddon, 1997), the authors propose that the correct dimensions and geometry of a nucleotide and not its hydrogen bonds are primarily responsible for the specific and efficient incorporation by polymerases. A comparison of Kool's reported 40-fold reduction in primer extension accuracy (dFTP vs. TTP opposite dA, (Moran et al., 1997a)) for the difluorotoluene nucleotides with the twofold drop in accuracy for the incorporation of **20** (ψTTP vs. TTP opposite dA) does not support their hypothesis. Instead, it may be that the reduced accuracy, observed with C-glycosidic analogs such as pseudothymidine, is also encountered by the C-glycosidic difluoro nucleotide. The additional 20-fold drop for dF may then be ascribed to the lack of hydrogen bonding capacity. Obviously, this proposal would require further investigation. It could be addressed by preparing a series of structural intermediates between thymidine and difluorotoluene to study and precisely quantify the various structural and non-covalent effects of unnatural nucleotides on DNA polymerases.

Finally, Breaker's argument that nucleic acid performs badly as a catalyst because it lacks structural rather than functional diversity, could be put to the test (Breaker, 1997b). Substituting thymidine with pseudothymidine creates an oligonucleotide that may have improved ability to form a greater diversity of tertiary structures, ultimately leading to improved catalysts. It would be of interest to perform in vitro selection on a DNA pool where all thymidines are replaced by ψT's to see how the substitution affects the resulting sequence motifs and physicochemical properties of the deoxyribozymes.

CHAPTER 5
TOWARDS DIRECTED EVOLUTION OF DNA POLYMERASES THAT RECOGNIZE AN EXPANDED GENETIC ALPHABET

Introduction

Preliminary screening experiments have identified two polymerases possessing the ability to incorporate the pyDAD and the puADA nucleotides (Lutz, 1997). The first, HIV type-1 reverse transcriptase mutant Y188L, was further investigated as reported in Chapter 3. The second, the wild-type of the DNA-dependent DNA polymerase from *Thermococcus* sp. 9°N-7 (Southworth et al., 1996) was particularly interesting because the single subunit enzyme was thermostable. However, in the absence of any structural information beyond the primary amino acid sequence, a project to generate a series of site-specific mutants, based solely on molecular modeling seemed not very promising.

A new technique, called directed molecular evolution, to create enzymes with novel properties has now changed the situation. The most attractive feature of this technique is that, in contrast to rational engineering, it requires no knowledge of the relationship between the amino acid sequence, structure, and mechanism of catalysis. A new project, dedicated to the directed evolution of 9°N DNA polymerase to obtain an enzyme that could incorporate the building blocks of an expanded genetic alphabet, was therefore initiated.

Directed Evolution of Enzymes

In Nature, evolution provides a mechanism for adapting and improving proteins for particular biological functions. Following Darwin's model of natural selection, a process of random mutagenesis and selection of the fittest proteins has generated an incredible diversity of biological function.

The same basic idea stands behind directed evolution, a new combinatorial approach that has recently been described by several groups (Arnold, 1996; Skandalis et al., 1997; Stemmer, 1994b). The method is based on the creation of enzyme libraries by random mutagenesis. While perhaps most of the mutations may destroy the enzyme, some of the mutations should also be beneficial to the enzyme when exposed to certain selection pressures. Some enzyme variants will have a higher catalytic turnover, while others may develop higher thermostability than the precursor. Applying an efficient assay system, the improved biocatalysts can be identified and isolated. These are either characterized or more frequently, serve as the template for another round of randomization and selection. This process can be repeated as many times as necessary until candidates showing the desired performance are found.

The successful application of the method is well documented and has been used not only to modify the substrate specificity of enzymes (Crameri et al., 1996; Hasson et al., 1998; Liu et al., 1997; Zhang et al., 1997), but also their physicochemical properties such as tolerance to organic solvents and thermostability (Giver et al., 1998; Moore & Arnold, 1996; Proba et al., 1998; Zhao & Arnold, 1999).

The method also has its prerequisites, however. The three cornerstones of every successful directed evolution experiment are: i) an effective mutation strategy, ii) a functional expression of the protein in a suitable microbial host, and iii) a fast and reliable assay system

for the identification of enzymes variants with the desired properties out of a pool of 10^4 to 10^6 members of the initial library.

Mutation Strategies

The first step in each round of directed evolution is the creation of an enzyme library. Such a pool of homologous sequences can be generated either asexually by random mutagenesis of a single parental sequence or by sexual evolution, using a number of homologous parental sequences (Arnold, 1998). While both approaches have been used successfully to improve an enzyme's performance, asexual evolution in the test tube tends to lead to the accumulation of deleterious mutations, thus limiting the level of improvement that can be achieved by this method. In contrast, sexual evolution enables both the quicker accumulation of beneficial mutations, and the removal of deleterious ones (Arnold, 1998).

For that reason, most recent experiments still start from a single target sequence, which is converted into a gene library by error-prone PCR (Cadwell & Joyce, 1994) (Figure 5-1A). For all the following cycles, however, it is not merely the best performing sequence, but the best ten or twenty sequences that are used for random mutagenesis, the first step in DNA shuffling. A mixture of all the different gene libraries is partially digested by DNase I treatment (Figure 5-1B). The fragmentation allows recombination of mutations in a sequence from one parental strand with mutations from other sequences. This takes place during the fragments reassembly by several rounds of self-primed PCR (Figure 5-1C). During reassembly of the gene library, mutations on the various fragments are recombined randomly. This crossover gives potentially beneficial mutations on separate fragments a chance to be combined, resulting in an overall improved enzyme. By using a second PCR in the presence of

primers, the fragments having the orginal size are recovered from the reassembly mixture (Figure 5-1D) (Stemmer, 1994a; Stemmer, 1994b; Zhao & Arnold, 1997b).

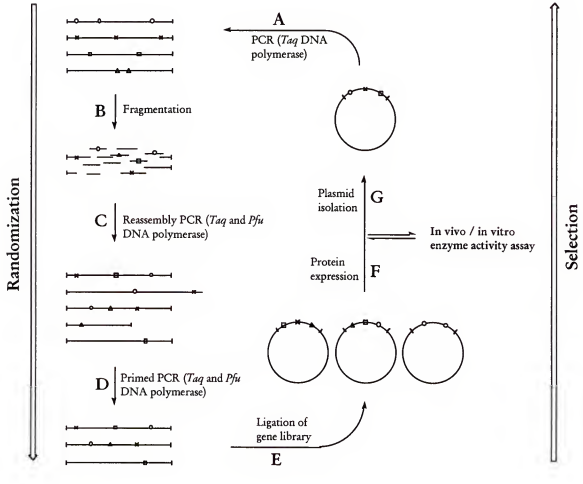


Figure 5-1. Working scheme for the directed evolution of enzymes. A) The experiment starts with PCR of the target gene. For PCR products >2 kb's, the natural error-frequency of *Taq* DNA polymerase introduces on average 4 mutations per gene (symbolized by O, X, and Δ). B) The PCR product is fragmented into 20-100 bp pieces by partial DNase I digestion. C) Fragments are reassembled by self-primed PCR. Using a mixture of *Taq* and *Pfu* DNA polymerase minimizes additional mutagenesis. D) Primed PCR of the extended fragments with *Taq/Pfu* DNA polymerase mixture amplifies correctly reassembled sequences. E) Digestion of the gene library by restriction enzymes, followed by ligation into the expression vector. F) The plasmid library is transformed into an expression host and protein is expressed in single-colony cultures. Improved enzymes are identified in vivo (complementation of auxotrophic hosts, phenotype) or in vitro (enzyme assays after protein isolation). G) The plasmid DNA of enzymes that show improved activity are isolated and used as starting sequences for the next round of randomization.

Mutation strategy for DNA polymerases.

For our DNA shuffling experiments, the gene sequence of 9°N-7 (exo^{DID}) DNA polymerase served as the basic template. The enzyme, a hyperthermostable DNA-dependent DNA polymerase, was isolated from a *Thermococcus* sp. found near a hydrothermal vent in the Gulf of California (Southworth et al., 1996). The initial enzyme library of 9°N DNA polymerase gene was generated by random mutagenesis with *Taq* DNA polymerase.

Alternatively, a pool of natural related sequences could be used to create a library of DNA polymerase sequences. Phylogenetic analysis identified four other thermostable DNA polymerases, the DNA polymerase from *Thermococcus literalis* (Vent®) (Kong et al., 1993; Perler et al., 1992), *Pyrococcus furiosus* (Fiala & Stetter, 1986; Lundberg et al., 1991), *Pyrococcus* sp. GB-D (Deep Vent®) (Jannasch et al., 1992), and *Pyrococcus* sp. KOD1 (Takagi et al., 1997), with at least 90% amino acid sequence similarity to 9°N DNA polymerase (Perler et al., 1996).

Regardless of whether we use only 9°N DNA polymerase, or a pool of related sequences, all polymerases must be exonuclease-deficient. The presence of exonuclease activity has previously been shown to cause problems in primer elongation experiments with nonstandard base pairs (Chapter 4). Pausing of the polymerase either at the position before a nonstandard nucleotide or shortly after formation of the nonstandard base pair has been unavoidable. At this point, the polymerase-primer-template complex starts to idle whereupon all dNTPs are turned into dNMPs as a result of repeated incorporation and exonucleolytic excision by the stalled polymerase (Derbyshire et al., 1995). Our work with DNA polymerases therefore focuses upon exonuclease-deficient versions. While the corresponding clones for the exo^- form of 9°N and Vent DNA polymerase were available from New England Biolabs, the exo^- mutant of *Pfu* DNA polymerase needed to be constructed.

Exonuclease-deficient DNA polymerases are prepared by mutation of the aspartate and the glutamate which are responsible for metal binding in the conserved exonuclease active site (Derbyshire et al., 1995). The 3'-5'-exonuclease activity is located in a discrete structural domain in the N-terminal region of the protein (Steitz, 1987). The replacement of the two metal ligands, D-355 and E-357 (in Klenow fragment, which corresponds to D-141 and E-143 in *Pfu*, Vent, or 9°N DNA polymerase) by alanines resulted in the complete disappearance of the exonuclease activity. This methodology was applied by New England Biolabs to create the exonuclease-deficient versions of Vent®, Deep Vent®, and 9°N-7 (Kong et al., 1993; Southworth et al., 1996). Although not described in much detail, Stratagene's *Pfu* (exo-) DNA polymerase presumably was created by the same mutations (Mathur, 1996).

Functional Protein Expression and Microbial Hosts

Following DNA shuffling, the PCR amplified product must be ligated back into a plasmid system and transformed into a suitable host organism for protein expression. Most of the early work on overexpression and isolation of thermostable DNA polymerases was done on *Taq* DNA polymerase (Desai & Pfaffle, 1995; Engelke et al., 1990; Pluthero, 1993), using pUC18-based vectors and DH1, DH5α, or INVαF' as the expression host. While this system seems to work well for *Taq* DNA polymerase, the leaky expression system for hyperthermostable DNA polymerases was reported to cause plasmid instability and growth inhibition as a result of cytotoxicity (Dabrowski & Kur, 1998; Lu & Erickson, 1997).

The BL21(DE3)pLysS/pET expression system

The most common expression system for thermostable DNA polymerases from hyperthermophiles is a combination of a pET vector with the *E. coli* B-strain BL21(DE3) or

BL21(DE3)pLysS (Dabrowski & Kur, 1998; Kong et al., 1993; Lu & Erickson, 1997; Niehaus et al., 1997; Southworth et al., 1996; Takagi et al., 1997). The pET expression system, first introduced by Studier et al. (Studier et al., 1990), uses a T7 promoter to control expression of a recombinant target protein. The highly specific promoter cannot be read by *E. coli* RNA polymerase, but requires T7 RNA polymerase to be transcribed. BL21(DE3) contains a chromosomal copy of the T7 RNA polymerase gene under control of a *lacUV5* promoter (Studier & Moffatt, 1986; Studier et al., 1990). Upon induction with IPTG, the expressed T7 RNA polymerase binds to the T7 promoter on the recombinant plasmid and transcribes the target DNA, located just downstream on the plasmid. For even tighter control of protein expression, BL21(DE3)pLysS carries the pLysS plasmid which encodes for T7 lysozyme. The lysozyme binds and inactivates the T7 RNA polymerase that is expressed at low levels due to leaking of the *lac* promoter (Studier et al., 1990). Upon induction with IPTG, the T7 RNA polymerase titrates out the inhibitor and goes on to transcribe the target protein.

Expression of archaea protein in *E. coli*

Beside plasmid instability and cytotoxicity, another problem for expression of archaeal proteins in *E. coli* has been reported. The redundant nature of the genetic code allows for multiple nucleotide triplets to encode for the same amino acid (Cubellis et al., 1990; Uemori et al., 1993). Rather than being used with equal frequency, certain codons are used preferentially by *E. coli* and other bacteria (Wade et al., 1992). Directly proportional with the species-dependent frequency of codon usage is the level of cognate tRNA (Yamao et al., 1991). Problems can arise when foreign genes are to be expressed in *E. coli*. Codons, such as AUA, AGA, and AGG are commonly found in archaeal and eukaryotic genes but rarely used by *E. coli* (Table 5-1) (Cubellis et al., 1990; Uemori et al., 1993). After induction of gene expression

in our pET system, large amounts of heterologous mRNA with multiple rare codons take over the translational machinery. This results in a transitory depletion of these rare tRNAs, forcing the translational machinery to pause, possibly causing early termination, misincorporation, and/or frameshifting (Brinkmann et al., 1989; Rosenberg et al., 1993). Extensive studies of the effects of tRNA depletion indicate that the position and clustering of these rare codons can have a dramatic effect on the performance of the ribosomal protein synthesis. In particular the presence of two or more consecutive rare codons in the 5'-region (as in archaeal DNA polymerases) disabled translation completely (Chen & Inouye, 1990; Chen & Inouye, 1994; Rosenberg et al., 1993).

Table 5-1. Codon usage comparison between *E. coli* and DNA polymerases from hyperthermophilic organisms.

Codons	AUA	AGA	AGG
E. coli	5.1	2.6	1.6
Pfu DNA polymerase	42.5	36.2	20.6
Pyrococcus sp. DNA polymerase	67.3	28.5	33.7
KOD DNA polymerase	38.7	16.8	28.4
Vent DNA polymerase	52.9	21.9	28.4
Deep Vent DNA polymerase	50.2	12.9	55.5
9°N DNA polymerase	32.2	3.87	36.1

Numbers are based on codons per one thousand. The entire genome for *E. coli* was considered while only the polymerase gene itself was counted for the hyperthermophiles. Data wee obtained from DISC (DNA Information and Stock Center, Tsukuba, Japan).

Two approaches have been published to overcome the limitations of protein expression by rare codons. Hu and coworkers have reported the systematic and specific replacement of multiple AGA/AGG codons with CGC and CGU, the arginine codons commonly used in *E. coli* (Hu et al., 1996). Their results showed that while the wild-type gene

was expressed only at low levels, the substituted sequence allowed the protein expression level to be raised up to 40-fold.

Alternatively, a plasmid carrying the *argU* gene which encodes for $tRNA^{AGA/AGG}$ can be co-transformed with the expression vector, carrying the target protein (Brinkmann et al., 1989; Jeong & Shin, 1998). Based on the same idea, pSBETa-c was constructed (Schenk et al., 1995). The vectors carry the T7 promoter region from pET3 in pSB161, a kanamycin-resistant vector that also contains the *argU* gene. In either case, the translation of thermostable DNA polymerase is supported by the co-transcription of $tRNA^{AGA/AGG}$.

Expression vector pAll17 and pSBETa

Two pET vectors, pAll17 and pSBETa (Figure 5-2) are used for our expression experiments. Both plasmids carry the identical T7 promoter region as part of the multiple cloning site (MCS). Using the NdeI and BamHI restriction sites, the shuffled PCR product is integrated unidirectionally. Both vectors contain modifications that make them particularly suited for the expression of archaea proteins. Additional termination sites upstream to the T7 promoter site in pAll17 actively reduce the transcription of heterologous mRNA and thereby prevent failure of translational machinery due to depletion of rare tRNA (Perler et al., 1992). In contrast, pSBETa co-transcribes the *argU* gene, which encodes for the rare *E. coli* $tRNA^{AGA/AGG}$ as previously discussed.

High-Throughput Screening Assays

Last but not least, directed evolution requires a fast and efficient assay to screen between 100 to 10,000 enzyme variants after each round of DNA shuffling for the desired

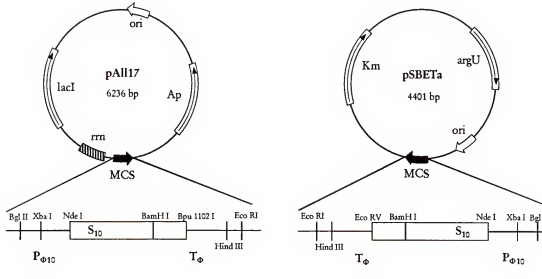


Figure 5-2. Expression vector pAll17 (left) and pSBETa (right). pAll17 is based on pET11c but contains four additional copies of transcriptional terminator (rrn) to modulate protein expression (Perler et al., 1992). pSBETa was constructed by ligation of the MCS fragment from pET3a (*Bgl*II/EcoRI) into pSB161, which already carries the *argU* gene for $tRNA^{AGA/AGG}$ (Schenk et al., 1995). The blow-up of the MCS shows the important restriction sites. $P_{\Phi 10}$ is the T7 promoter site, S_{10} indicates the translational start. The terminator sequence is indicated by T_{Φ}.

properties. As already discussed in Chapter 4 (High-throughput screening for DNA polymerases), an in vivo assay is not possible due to the elevated reaction temperature necessary to test the hyperthermophilic DNA polymerases and the inability to generate the triphosphates of pyDAD and puADA in vivo. However, the in vitro screening of DNA polymerases by primer extension and SPA has proven itself a valuable tool for enzyme screening under any desirable reaction conditions.

Materials and Methods

Experimental procedure

A summary of all chemicals and enzymes used is listed in Appendix B. The plasmids for 9°N-7 DNA polymerase (exo^{DID}) (pNEB916) (Southworth et al., 1996), as well as for Vent* (pAKK4) and Vent*(exo^{-}) (pALK1) (Kong et al., 1993; Perler et al., 1992) were obtained from New England Biolabs, Beverly MA. The pETpfu, carrying the *Pfu* DNA polymerase was a gift from H.P. Erickson, Duke University, Durham NC; (Lu & Erickson, 1997). The HIV type-1 reverse transcriptase mutant Y188L was prepared by Jeong Ho Park (for details see Chapter 3).

Primer and templates. Desalted standard oligonucleotides were ordered from Integrated DNA Technologies (Corelville MD) and used without further purification. The SPA template with a single pyDAD in position 25 (marked k) was prepared as described in Chapter 2 (Materials and Methods). Restriction sites are underlined (NdeI, BamHI, EcoRI, PstI). Gene sequences are printed in *italic*. The region with the restriction sites was kept constant for all polymerases. The length of the complementary strand was adjusted, based on the predicted melting temperature of the primer sequence. In primers for site-directed mutagenesis, the codon changes are highlighted in bold (P5-7, P5-8, P5-9, and P5-10).

9°N DNA Polymerase

P5-1: 5'-CTCTAGAATTCAGGAGGAAGTT*CATATGATTCTCGATACCGACTACATC*-3'

P5-2: 5'-AAGCTTCTGCAGGGATCC*TCACTTCTTCCCCTTCACCTTCAGCCA*-3'

P5-3: 5'-GAATTCAGGAGGAAGTT*CATATG*-3'

P5-4: 5'-AAGCTTCTGCAGGGATCCTCA-3'

Pfu DNA Polymerase

P5-5: 5'-GTCTAGAATTCAGGAGGAAGTTCA*TATGATTTTAGATGTGGATTAC*-3'

P5-6: 5'-GAAGCTTCTGCAGGGATCC*CTAGGATTTTTTAATGTTAAGCC*-3'

P5-7: 5'-AAGGTCAGGT**CTTT**CTCCCTGATAATCCTGAGAAATC-3'

P5-8: 5'-GGGAGAAAGACCCTGACATTATAGTTACTTATAATGG-3'

P5-9: 5'-GGTTGCTATAGCGAAGGCAAGAATCTTTAGCTCTTC-3'

P5-10: 5'-GCCTTCGCTATAGCAACCCTCTATCACGAAGGAGAAGAG-3'

Vent DNA polymerase

P5-11: 5'-GTCTAGAATTCAGGAGGAAGTTCA*TATGATACTGGACACTGATTACATAACA AAAGATG*-3'

P5-12: 5'-GAAGCTTCTGCAGGGATCC*CTACCTCTTGAGCCATGCATCTAAGC*-3'

Scintillation proximity assay

P5-S: 5'-GCGTAATACGACTCACTATAG-3'

T5-S1: 5'-TATGATCGTCTATAGTGAGTCGTATTACGC-3'

T5-S2: 5'-TATGkTCGTCTATAGTGAGTCGTATTACGC-3'

Construction of Pfu (exo⁻) DNA polymerases. *Removal of the BamHI site*: Two 100 μL-reaction mixtures in *Taq* reaction buffer (Promega), each containing pETpfu (~50 ng), dNTPs (200 μM each), magnesium chloride (1.25 mM), and either primers P5-5/P5-7 (1 μM each, 5'-fragment) or P5-6/P5-8 (1 μM each, 3'-fragment) were incubated with 2.5 U of *Taq* DNA polymerase and 1.25 U *Pfu* DNA polymerase. Prior to cycling, the mixtures were heated to 95°C for 5 min, followed by 35 cycles of amplification (94°C, 45 sec; 55°C, 45 sec; 72°C, 3 min). The PCR products were gel-purified (Appendix C5) and reassembled in a second

PCR. In a 50 μL-reaction mixture in *Taq* reaction buffer, approximately 20 ng of each fragment were mixed with P5-5 and P5-6 (1 μM each), dNTPs (200 μM each), magnesium chloride (1.25 mM), and 2.5 U *Taq* DNA polymerase and 1.25 U *Pfu* DNA polymerase. After the initial heat denaturing (94°C, 5 min), the sample was amplified over 35 cycles (94°C, 45 sec; 55°C, 45 sec; 72°C, 3 min). The gel-purified PCR product was ligated into pCR®-XL TOPO (Invitrogen) and transformed into *E. coli* TOP10. After overnight incubation on LB/Kan plates (Appendix B5), colonies were picked and grown overnight in liquid LB medium (Appendix C2). The plasmid DNA was isolated (Appendix C3) and the mutations checked by restriction analysis with NdeI/BamHI (Appendix C7) and DNA sequencing of the corresponding region.

Removal of the exonuclease activity: Using one of the sequenced (BamHI) plasmids, another two reaction mixtures with ~50 ng template and primer P5-5/5-9 (1 μM each) or P5-6/P5-10 (1 μM each) were prepared. Keeping all the concentrations and conditions as previously described for the BamHI removal, the two-stage PCR was performed and the reassembled and purified material ligated into pCR®-XL TOPO. After transformation into *E. coli* TOP10, plasmid DNA was isolated from overnight cultures of randomly picked colonies in LB/Kan. Restriction analysis and DNA sequencing confirmed the successful mutagenesis. The polymerase activity of the exo⁻ mutant was confirmed by restriction digestion of the *Pfu* (exo⁻) gene with NdeI/BamHI, followed by ligation into pAll17 or pSBETa and protein expression in *E. coli* BL21(DE3)pLysS (for detailed protocol see "Ligation and cloning of DNA libraries" and "Protein expression of polymerase libraries").

DNA shuffling. The gene for 9°N DNA polymerase was amplified by PCR from pNEB916. In a total reaction volume of 100 μL, pNEB916 (45 ng) was incubated with primer P5-1 and P5-2 (1 μM each), as well as dNTPs (200 μM each), magnesium chloride (1.25 mM),

and *Taq* polymerase buffer (10x). The hot-start PCR with *Taq* DNA polymerase (40 U) was cycled (94°C, 1 min; 55°C, 1 min; 72°C, 3 min) over 30 rounds with a 10 min polishing at the end. The product was purified by agarose gel electrophoresis (see Appendix C5).

The purified 9°N gene (~8 μg) was dissolved in Tris-HCl (50 mM, pH 7.5) and manganese chloride (10 mM). The reaction mixture (200 μL) was then incubated for at least 15 min at 15°C prior to addition of DNase I (0.4 μL, freshly diluted with Tris-HCl (50 mM, pH 7.5) to 1 U/μL). In one-minute intervals, 50 μL aliquots were quenched in 300 μL QIAEX II buffer, containing EDTA (2 μL, 0.5 M, pH 8). The DNA fragments were isolated, using the QIAEX II DNA purification system and loaded on an agarose gel (1.4%). Fragments, 20-100 bp in length, were recovered with the QIAEX II system.

The fragments were reassembled by a self-priming PCR. In 100 μL reaction volume, DNA fragments (1.25 μg) and dNTPs (200 μM) were mixed with *Pfu* polymerase buffer (10x). After an initial denaturing step (94°C, 30 sec), cloned *Pfu* DNA polymerase (5 U, 2.5 U/μL) was added, and the tube cycled (94°C, 1 min; 55°C, 1 min; 72°C, 7 min) over 40 rounds with a 10 min polishing at the end of the last cycle.

An aliquot of the self-priming PCR was then used for the primed PCR. Between 2-20 μL of the previous PCR were mixed with primer P5-3 and P5-4 (1 μM), dNTP (200 μM), and magnesium chloride (1.25 mM) in *Taq* polymerase buffer (10x). The mixture (30 μL) was denatured (94°C, 30 sec), *Taq* DNA polymerase (1.25 U, 5 U/μL) and cloned *Pfu* DNA polymerase (1.25 U, 2.5 U/μL) were added, and cycled (94°C, 1 min; 55°C, 1 min; 72°C, 5 min) over 35 rounds with 10 min polishing at the end of the last cycle. The shuffled full-length product was isolated by agarose gel electrophoresis (Appendix C5).

Ligation and cloning of DNA libraries. In two separate reactions mixtures of 30 μL in BamHI buffer (supplied by vendor), the shuffled fragment (~4 μg) and the expression

vector pSBETa (~4 μg) were incubated with NdeI and BamHI (1 μL each) overnight at 37°C. The DNA fragments were purified by agarose gel electrophoresis (Appendix C5), followed by ethanol precipitation. After re-equilibrating, the NdeI/BamHI fragment of pSBETa (~50 ng) and the digested polymerase insert (~80 ng) were mixed in T4 Ligation buffer (supplied by vendor) and the volume adjusted to 10 μL with water. After addition of T4 DNA ligase (1 U, 0.1 μL), the reaction was incubated at 25°C for 5-6 h. The ligation was quenched by dilution with 80 μL water, followed by desalting of the product (Appendix C4). After lyophilization, the DNA was redissolved in TE buffer (50 μL) and aliquots (5 μL) were transformed into electrocompetent *E. coli* TOP10. The ligation efficiency was determined by plating an aliquot on LB/Kan plates and picking single colonies the next day. After overnight growth of the individual colonies, plasmid prep followed by restriction analysis with NdeI/BamHI identified the percentage of successful ligations. For protein expression, an aliquot (25 μL) from the transformation was incubated overnight in LB/Kan medium (50 mL). The plasmid DNA (pSBETa-9°N library) was isolated from the mixed culture by Wizard MidiPrep. Libraries of 9°N in pAll17 were prepared the same way but Kan was replaced by Amp.

Protein expression of polymerase libraries. pSBETa-9°N library (~100 ng) was chemically transformed into BL21(DE3)pLysS (50 μL aliquots). Following the manufacturer's protocol, aliquots (50 μL) were plated on LB/Kan/CM plates and incubated overnight.

For the preculture, single colonies were picked and incubated individually in microtiterplate wells, containing 180 μL LB/Kan/CM medium. The plates were grown at 37°C overnight and were stored at 4°C (motherplate). Protein expression was set up in microtiterplates or 10 mL Falcon tubes. Aliquots from the preculture (2.5 μL (for microtiterplates)/ 10 μL (for Falcon tubes)) were used to inoculate the expression cultures in LB/Kan/CM medium (180 μL/ 2 mL), followed by incubation at 37°C. When the cultures

reach log-phase (OD_{600}: 0.1-0.3/ 0.4-0.7), protein expression was induced by addition of IPTG (0.8 μL/8 μL) to a final concentration of 0.4 mM. Cells were harvested 2-3 h after induction by centrifugation or filtration. Cell pellets were resuspended in TN-buffer (20 μL) and stored at -80°C until further used.

To isolate the polymerase, the resuspended cells were lysed by heating (96°C, 15 min). After the heat treatment, all samples were stored on ice for 10-20 min, followed by centrifugation (10,000 *g*, 5 min). The clear supernatant was used directly for primer extension and SPA.

SPA for DNA polymerase libraries. *Primer/Template preparation:* 5'-O-Biotinylated primer (P5-S) and template (T5-S1 or T5-S2) were mixed in a 1:7 molar ratio in TE-buffer and adjusted to a final primer concentration of 10 pmol/μL. The solution was heat-denatured (95°C, 2 min), followed by slow annealing over 1 h. The annealed primer/template complex can be stored at -20°C for at least 6 months.

SPA reaction mixture: (amounts per microtiterplate - 100 reactions) Annealed primer/template complex (10 μL) was mixed with SPA Pol buffer (500 μL, Appendix B5), magnesium sulfate (100 μL, 1 M), dGTP (375 μL, 1 μM), dATP (375 μL, 1 μM), dCTP (375 μL, 1 μM), TTP (125 μL, 1 μM), [^{3}H]-TTP (30 μL, 90-130 Ci/mmol), and 1.11 mL water. For assays with nonstandard nucleotides, dXTP (375 μL, 1 μM) was added (the extra volume was compensated for by reducing the water). Aliquots (30 μL) were pipetted into 96-well PCR reaction plates and stored on ice. The SPA reaction mixture can be stored on ice for up to 1 h.

SPA: On ice, the clear supernatant from the polymerase expression (20 μL) was mixed with the prepared SPA reaction mixture. The plates were incubated at 72°C. After 60 min, the reactions were quenched by addition of EDTA (20 μL, 0.5 M, pH 8), and aliquots (60 μL) were transferred into streptavidin-coated Flashplates (NEN Life Sciences). To allow binding of the

primer/template complex, the plates were stored at RT for 20-30 min before being washed with Tris-HCl (200 μL, 20 mM, pH 8). After removing the liquid phase from the wells, fresh Tris-HCl (260 μL, 20 mM, pH 8) was added to each well, the plate was sealed and counted in a TopCount® LSC (Packard Instruments).

Plasmid isolation - identification of mutants. Based on the data from the SPA, an aliquot (10 μL) of the corresponding colony on the motherplate was incubated in LB (Kan/CM) medium (2 mL). The overnight culture was centrifuged and the plasmid DNA isolated (Appendix C3). The material served as the template for the next round of random mutagenesis or was submitted for DNA sequencing.

Results and Discussion

A general protocol for directed evolution of DNA polymerases that recognize non-standard base pairs was developed and tested on the 9°N-7 (exo^{DID}) DNA polymerase. Gene libraries of the randomly mutagenized and shuffled DNA polymerase were prepared and transferred into *E. coli*. Individual colonies were cultured, protein expression induced, and the resulting enzymes tested by primer extension and SPA high-throughput screening. To demonstrate the generality of the methodology, libraries of exonuclease-deficient versions of *Pfu* and Vent DNA polymerase were prepared by the same procedure.

Exonuclease-Deficient *Pfu* DNA Polymerase by Overlapping PCR

An exonuclease-deficient version of *Pfu* DNA polymerase was prepared by site-directed mutagenesis on two residues, D141A and E143A. In addition to the exonuclease

activity, an internal BamHI restriction site was removed by site-directed mutagenesis. Identified in preliminary cloning experiments, the internal BamHI sequence would make it impossible to use our universal restriction sites (NdeI/BamHI) for DNA shuffling. The exonuclease active site, as well as the BamHI restriction site was mutated in two separate steps, using overlapping primer extension as shown in Figure 5-3.

Removal of the BamHI site. Based on the DNA sequence of *Pfu* DNA polymerase (Genbank accession No. D12983), primer P5-7 and P5-8 were synthesized. The BamHI-site was removed by introducing two silent mutations of K201 (AAG to AAA) and D202 (GAT to GAC). The success of the mutagenesis was checked by restriction analysis with NdeI/BamHI. Removal of the internal BamHI site was successful in six out of ten plasmids. The previously observed fragment around 0.6 kb disappeared and the pol gene fragment showed the expected size around 2.3 kb. DNA sequencing of the region further confirmed the codon changes (Figure 5-4).

Removal of the exonuclease activity. Exonuclease-deficient *Pfu* DNA polymerase was constructed by two expressed mutations at codon D141 and E143. Both positions were mutated into alanines by codon switches at D141 (GAT to GCT) and E143 (GAA to GCA). The restriction analysis of the reassembled and cloned *Pfu* (exo') gene indicated that none of the colonies accidentally carried the wild-type gene with the internal BamHI site. DNA sequencing confirmed the successful mutagenesis of the two positions into alanines (Figure 5-5). After subcloning into the pAll17 or pSBETa expression vectors, the catalytic activity of *Pfu* (exo') DNA polymerase was confirmed by primer extension experiments and SPA analysis.

In summary, we applied site-directed mutagenesis in two locations of the *Pfu* DNA polymerase gene to remove not only the 3'-5'-exonuclease activity but also an internal BamHI restriction site. The success of the procedure was demonstrated by DNA sequencing, as well as

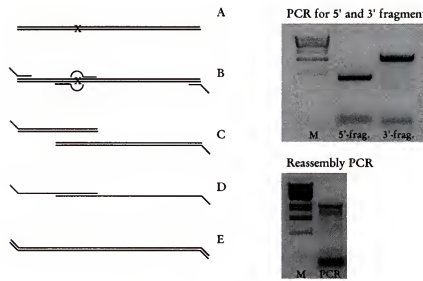


Figure 5-3. Site-directed mutagenesis by overlapping primer extension. A) The position of the desired mutation in the gene is marked with X. B) In two separate reactions, the internal primers, carrying the mutation, are PCR amplified with the external primers. C) PCR generates a 3'- and a 5'-fragment (top picture). D) The two fragments are annealed and PCR amplified. E) The reassembled PCR product with the correct restriction sites can be digested and ligated (bottom picture).

```
                                  201  202
Pfu (wt)                 …tatcagggag aag gat cctg acattatagt…

Pfu (B⁻)#1               …TATCAGGGAG AAA GAC CCTG ACATTATAGT…
Pfu (B⁻)#2               …TATCAGGGAG AAA GAC CCTG ACATTATAGT…
Pfu (B⁻)#3               …TATCAGGGAG AAA GAC CCTG ACATTATAGT…
                                  lys  asp
```

Figure 5-4. Site-directed mutagenesis of wild-type *Pfu* DNA polymerase gene by overlapping primer extension. The BamHI restriction site was removed by introduction of two silent mutations at codon 201 and 202. DNA sequencing of three BamHI⁻ sequences, selected after restriction enzyme digestion had indicated the absence of the internal BamHI site, confirmed the site-specific mutagenesis at the two positions.

by detection of polymerase activity of the expressed *Pfu* (exo⁻) DNA polymerase. To complete the project, an additional exonuclease activity assay may be performed (Southworth et al.,

1996). In principle, the *Pfu* gene can now be used, together with Vent and 9°N, for future DNA family shuffling (Crameri et al., 1998).

DNA Shuffling

DNA shuffling can be divided in four steps: i) random mutagenesis of the gene to be shuffled, ii) fragmentation of the gene pool with DNase I, iii) reassembly of the fragments by self-primed PCR, and iv) amplification of the shuffled gene pool from the reassembly mixture. A gel picture, summarizing the entire shuffling procedure, is shown in Figure 5-6.

```
                                141  142  143
Pfu (wt)              …tcttgccttc gat ata gaa a ccctctatca…

Pfu (B⁻/e⁻)#1         …TCTTGCCTTC GCT ATA GCA A CCCTCTATCA…
Pfu (B⁻/e⁻)#2         …TCTTGCCTTC GCT ATA GCA A CCCTCTATCA…
                                  asp->   glu->
                                  ala     ala
```

Figure 5-5. Exonuclease-deficient mutant of *Pfu* DNA polymerase (BamHI'), created by overlapping primer extension. After ligation and transformation into *E. coli*, plasmid DNA from two randomly picked colonies was isolated and the two mutations confirmed by DNA sequencing. The two substitutions at D141A and E143A remove the metal-binding ligands in the exonuclease's active site (Kong et al., 1993).

Random mutagenesis. The initial pool of randomly mutated polymerase genes was prepared by PCR amplification using *Taq* DNA polymerase. Sequences <2 kb usually require the addition of manganese to increase the error frequency of *Taq* to the desired 2-6 mutations per gene (Cadwell & Joyce, 1994). For the 2.3 kb polymerase gene however, the natural mutation rate of $\sim 5 \times 10^{-5}$ (Barnes, 1992; Lundberg et al., 1991) is sufficiently high to introduce on average three nucleotide changes per sequence. DNA sequencing of plasmids,

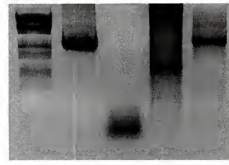


Figure 5-6. Gel picture of the four stages in the DNA shuffling protocol. M) Size marker: λ_{DNA} (*Bst*EII digest), i) Large-scale random-mutagenesis PCR of the 2.3 kb 9°N DNA polymerase gene, using *Taq* DNA polymerase, ii) Digestion of the PCR product into 50-200 bp fragments with DNase I, iii) Reassembly mixture after self-primed PCR with *Pfu* DNA polymerase, iv) regular, primed PCR amplification of the reassembly mixture with *Taq*/*Pfu* DNA polymerase mixture, recovering a band of the original size.

containing the gene for 9°N DNA polymerase, from ten randomly picked colonies after mutagenesis with *Taq* DNA polymerase and shuffling confirmed the appearance of 2-4 nucleotide changes per sequence. Approximately 50% of the sequence of all ten genes were sequenced (700 bp from either side). From the 12 mutations identified, transitions from adenosine to guanosine and thymidine to cytidine appeared four times, while in two cases, guanosine was exchanged for adenosine. Only a single transversion from adenosine to cytidine and one deletion were detected. Generally, a substitution of AT for GC base pairs was observed, a shift that had previously been reported for unbiased dNTP pools (Cadwell & Joyce, 1994). Its impact on the fitness of the gene pool is not clear but it is striking that most improved enzymes from in vitro evolution experiments show a unusually high number of AT to GC shifts (Crameri et al., 1996; Moore & Arnold, 1996; Zhao & Arnold, 1997a; Zhao & Arnold, 1999). If necessary, the drift can be controlled by adjusting the concentrations of the individual dNTPs in the PCR (biased dNTP pool).

DNase I fragmentation. Fragments of the gene pool, 50-200 bp in size, were generated by DNase I treatment (Figure 5-6). In the presence of manganese ions, DNase I cuts double-stranded DNA at random (Ausubel et al., 1997; Lorimer & Pastan, 1995). The degree of digestion can be controlled by the reaction time. Under the reported experimental conditions (see Materials and Methods), agarose gel electrophoresis indicated the complete conversion of the 2.3 kb gene into 50-200 bp fragments within 2-3 minutes. This fragment size-range was used for the reassembly reaction. When further digested to 20-50 bp fragments, the mixture could not be reassembled, probably due to the short overlapping regions, which failed to reanneal under the conditions used for self-primed PCR.

Self-primed reassembly PCR. The small fragments were recombined by multiple cycles of self-annealing, primer extension, and denaturation. Beside the fragment's size, discussed in the previous paragraph, the quantity of fragments proved to be crucial to the successful reassembly of gene size fragments. Only when the amount of the fragment pool was above the experimentally determined threshold of 1 μg DNA, could we recover our original gene. We assume that at lower fragment concentrations certain pieces of the full-length product are absent, preventing reassembly.

Primed PCR amplification. From the pool of shuffled and reassembled sequences, the correctly reassembled target genes was amplified by a regular primed PCR. As in the reassembly step, the quantity of mixture proved important for the success of the experiment. A series of PCRs, using various amounts of reassembly mixture, was usually performed (Figure 5-7).

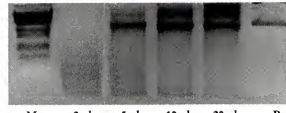


Figure 5-7. Primed PCR amplification with various amounts of reassembled material. To small amounts of template material (2 μL) result in failure to amplify the correct-size band while excess of template (> 10 μL) produces increasing amounts of extended PCR product. Reference (R) is the original size gene for 9°N DNA polymerase, M is the marker λ_{DNA}(*Bst*EII digest).

Protein Expression System for DNA Polymerases

The complete protein expression system is divided into three steps: i) ligation of the PCR product into a vector, ii) expression of the heterologous protein in *E. coli*, and iii) isolation and purification of the target protein.

Ligation of the gene library

Several techniques to ligate PCR products into an expression vector were tested. The rapid integration by TA cloning, using the TOPO XL PCR cloning kit specially designed for long PCR products worked very efficiently. Nevertheless, the resulting construct could not function as an expression vector and additional subcloning resulted in the undesired "dilution" of the gene library.

The direct ligation of the NdeI/BamHI-predigested PCR product and expression vector by T4 DNA ligase gave variable results. First, the gel purification of the fragments prior to the ligation was necessary to prevent self-ligation of the vector or the much faster ligation of the short 5' and 3'-ends from the PCR product digestion. Second, the efficiency of the ligation

dropped with increasing size of the expression vector. The restriction analysis for pUC18, pBCSK+, and pSBETa (all ~4 kb) indicated ligation efficiencies around 80-90%, while for pAll17 (~6.2 kb) the number of successful recombinants dropped to 50%. The libraries created with the shuffled 9°N-gene pool were named pUC18-9°N, pAll17-9°N, and pSBET-9°N.

Expression vector and host organism

The expression of 9°N DNA polymerase was tested in different host/vector systems. Beside the traditional pUC18- and pUC19-based expression systems in *E. coli* K-12 strains, *E. coli* BL21(DE3)pLysS was transformed with pAll17 and pSBETa.

***E. coli* K-12 strains as host for pUC18 or pBCSK+.** Initial cloning and expression experiments with 9°N DNA polymerase were performed with pUC18 in various K-12 strains. While only *E. coli* HB101 expressed the DNA polymerase at detectable levels (Figure 5-8), the system was found to be extremely unstable. Problems arose from the very slow expression of target protein, requiring extended induction periods of up to 20 hours. Studier et al. (Studier et al., 1990) have reported that ampicillin is already completely degraded when the cultures start to look turbid. Consequently, an extended incubation time favors cells lacking plasmid to overgrow the culture. Replacement of the pUC18 vector by pBCSK+, a pUC19-based vector with chloramphenicol resistance did not result in detectable amount of DNA polymerase after induction with IPTG. Additional concerns about the cytotoxicity of hyperthermostable DNA polymerases (Dabrowski & Kur, 1998) on the host organism and the known leakiness of the *lac* promoter lead to a switch to the tighter-controlled pET® expression system.

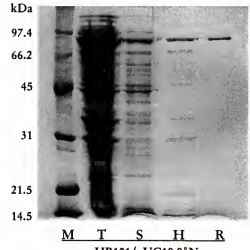


Figure 5-8. pUC18-based expression of 9°N DNA polymerase in *E. coli* HB101 (overnight induction with 0.5 mM IPTG at 37°C). Next to the size marker (M), a total protein sample (T) (1 mL culture pellet in 50 μL PAGE gel loading buffer, heat-lysed (96°C, 15 min) and centrifuged), a soluble protein sample (S) (20 mL culture pellet in 2 ml Tris-HCl (50 mM, pH 8) sonicated 3x 20 sec on ice, centrifuged, and mixed with PAGE gel loading buffer), and heat-lysed product (H) (2 mL culture pellet in 2 mL TN-buffer heat-lysed (96°C, 15 min) and clear supernatant mixed with PAGE gel loading buffer) were run with a reference sample of 9°N DNA polymerase (R). Enzyme activity was confirmed in a separate primer extension assay.

E. coli B-strain as host for pAll17 or pSBETa. The *E. coli* B-strain BL21(DE3)pLysS was tested for the expression of thermostable DNA polymerases. Following transformation with pAll17-9°N, single colonies were cultured and expression of 9°N DNA polymerase was monitored by SDS-PAGE (Figure 5-9, left). Within 3 h of induction with IPTG, the polymerase band was clearly visible. Furthermore, the plasmid stability was improved; cultures could be regrown multiple times over 24 hours without lost of the plasmid.

The problem with extended incubation times still remained, however. As with the HB101/pUC18 system, the expression of significant amounts of protein still required 6-16 h incubation after induction with IPTG. In addition, protein expression and total activity in cell

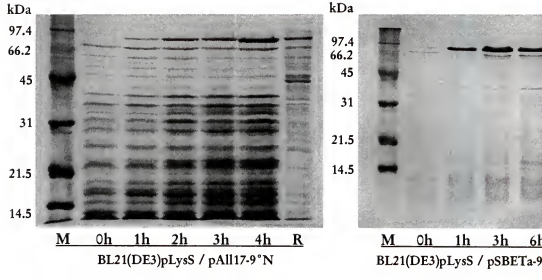


Figure 5-9. Protein expression of 9°N DNA polymerase in BL21(DE3)pLysS, using pAll17 or pSBET as expression vector. Cultures were induced with IPTG at time t=0 and aliquots (0.6 mL for pAll17-9°N and 0.3 mL for pSBETa-9°N) were taken after the indicated times. A sample from the soluble protein fraction was analyzed by SDS-PAGE. The strong band that appears upon induction with IPTG runs slightly below the 96 kDa marker band and is identical in size (93 kDa) with a reference sample (R) of 9°N DNA polymerase. All samples were tested positive for DNA polymerase activity. The same results were obtained for Pfu (exo') DNA polymerase (data not shown).

cultures were still the highest at 30°C and modest IPTG concentrations of 0.4 mM, rather than the usual 37°C and 1-5 mM IPTG. It appears that despite the additional termination sites, the effects of rare codon usage in 9°N DNA polymerase continue to affect protein expression.

The significance of the rare codon usage on the overexpression of hyperthermostable DNA polymerases in *E. coli* was studied, using pSBETa (Schenk et al., 1995). Upon induction with IPTG, the plasmid not only expresses the target protein but also simultaneously co-transcribes the rare $tRNA^{AGA/AGG}$. The difference in protein production between pAll17 and pSBETa in E. coli BL21(DE3)pLysS was analyzed by SDS-PAGE (Figure 5-9).

The protein gels indicate that both vectors, pAll17 and pSBETa, lead to protein expression. While expression of DNA polymerase in pAll17 was still increasing after 4 h, the pSBET system had reached its maximum after only 3 h. The quicker response to IPTG induction by pSBET was reconfirmed in multiple experiments with 9°N and Pfu DNA polymerase. A quantitative comparison, based on the amounts of cell pellets used for either SDS-PAGE indicates about two to threefold more protein from the pSBETa expression vector (Figure 5-9). About twice the cell culture volume (at comparable cell densities) was used for the pAll17 sample than for the pSBETa one. Despite the larger sample amount, the band after 4 h (pAll17) is still slightly smaller than the product band after 3 h (pSBET). The purity of the preparation can be estimated from the SDS-PAGE and is clearly higher for the pSBETa sample.

A very important observation was made for all BL21(DE3)pLysS cultures. The pLysS-carrying strains could not be stored at 4°C over an extended time-period. The dead cells in the culture release their T7 lysozymes, normally safely contained in the cell's cytoplasm, into the culture medium, which causes cell hydrolysis within 1 or 2 days. However, the problem can be overcome by preparing glycerol cultures (Appendix C1).

Protein expression libraries

Based on the preliminary tests with the various hosts and expression vectors, BL21(DE3)pLysS, carrying the pSBETa-9°N library was chosen for larger scale expression and activity screening. Following the ligation of the PCR library into pSBETa, protein expression in B-strains required the preculturing of the plasmid library. In contrast to the K-12 strains where the plasmid can be transformed and expressed immediately, the efficiency for direct transformation of ligation product into BL21(DE3)pLysS was effectively zero. Instead, the

ligation mixture was transferred into E. coli TOP10 first, and grown in an overnight culture. An aliquot of plasmid DNA, recovered by plasmid prep was then successfully transformed into the BL21 host.

The chemical transformation of the pSBETa-9°N library into BL21(DE3)pLysS (50 μL aliquots, Invitrogen One-Shot cells) was performed with a large plasmid excess (100 ng). Accordingly the transformation efficiency was low ($\sim 3 \times 10^4$ transformant/μg), yielding 3000 to 3500 isolated colonies (6 x 50 μL aliquots were plated).

Single colonies were picked randomly and grown in round-bottom microtiterplates at 37°C. The cell densities showed a Gaussian distribution at all stages of growth, indicating that none of the colonies had lost the plasmid. Protein expression was induced at an average $OD_{595}^{(MTP)}$ of 0.15 ($\pm$0.05), which corresponds to a regular OD (1-cm pathlength) of 0.5. The colonies were harvested after 3 h at an $OD_{595}^{(MTP)}$ of 0.42 ($\pm$0.1).

Purification of overexpressed DNA polymerases

The overexpressed polymerase was isolated by heat lysis, followed by centrifugation. The procedure yielded polymerases of 70-90% purity as determined by SDS-PAGE (Figure 5-10). Taking advantage of the thermophilic nature of the target protein, the heat treatment denatures all host proteins while the thermophilic enzyme stays in solution. Concerns about destruction of the target enzyme under these conditions were addressed. In preliminary tests, purified 9°N DNA polymerase was incubated in TN-buffer for up to 30 min at 96°C without significant loss of activity (< 10%). Similar results were obtained for the standard addition of pure enzyme to *E. coli* cultures, followed by immediate heat treatment and activity assay. No evidence for accidental co-precipitation of the polymerase during the subsequent centrifugation

was found. At a later stage of the project, UV spectroscopy and ethidium bromide staining of the enzyme preparation revealed the presence of large amounts of nucleic acid in the sample.

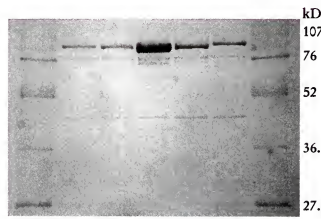


Figure 5-10. Example for SDS-PAGE analysis of five 9°N DNA polymerase colonies, cloned in pSBETa and expressed in BL21(DE3)pLysS. The 1-mL culture broths were harvested 3 h after induction. The suspension was centrifuged, the pellet resuspended in 10 μL TN-buffer and heat-lysed (96°C, 15 min). After cooling to RT and centrifugation, the supernatant was mixed with SDS gel loading buffer and analyzed (Appendix C9). Lanes with prestained SDS-PAGE low range standard (BioRad) are marked.

DNA Shuffling and Protein Expression of Vent and *Pfu* DNA Polymerases

Using the same protocol as reported for 9°N DNA polymerase, gene and plasmid libraries of the exonuclease-deficient versions of Vent and *Pfu* DNA polymerase were prepared. The overlapping region of the PCR primers for both genes, containing all the necessary unique restriction sites, were sequence-identical with the primers of the 9°N gene. Universal primer sites are necessary for family DNA shuffling (Crameri et al., 1998), another way to generate large libraries of DNA polymerase with the potential for expanded substrate specificity.

Ligation of both gene libraries into pSBETa worked with similar efficiency as for the 9°N gene. Protein expression was performed only with *Pfu* DNA polymerase. Polymerase activity was detected and SDS-PAGE analysis indicated approximately the same expression level as for 9°N DNA polymerase.

SPA with the pyDAD/puADA Base Pair

The SPA was used to screen libraries of thermostable DNA polymerases for enzymes that can specifically incorporate pyDAD and puADA in the presence of all the standard nucleotides. Encouraged by the successful search for a polymerase that can incorporate pseudothymidine (Chapter 4), a similar set of primer extension reactions was designed to identify new lead polymerases for pyDAD and puADA. In addition to the regular [^{3}H]-TTP incorporation, the utilization of [^{3}H]-dXTP by the polymerase was monitored.

Preliminary SPA with HIV reverse transcriptase (Y188L)

A series of preliminary studies on the performance of the SPA with pyDAD and puADA were conducted, using HIV-RT (Y188L). Similar to the primer extension reactions in Chapter 4, a set of seven reactions was run to monitor the incorporation of [^{3}H]-TTP and [^{3}H]-dXTP (Figure 5-11).

Beside the necessary positive controls and background reactions, this set of primer extension reactions allows us to identify polymerases that incorporate d(puADA)TP. In addition, the data from the experiments give us qualitative information, as to whether the polymerase stalls or elongates the primer beyond the nonstandard base pair.

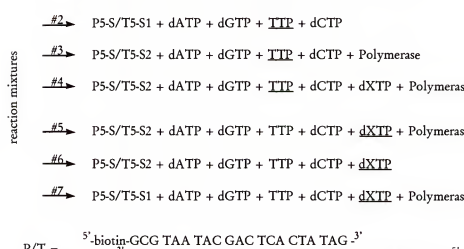


Figure 5-11. Preliminary SPA tests with [^{3}H]-TTP and [^{3}H]-dXTP with HIV-RT (Y188L). T5-S1, the template in reaction 3, 4, and 7 contains only standard oligonucleotides (k=A). T5-S2, the templates in reaction 1, 2, 5 and 6 carries pyDAD at the position k. The underlined dNTP was tritium-labelled and acted as a reporter for successful incorporation. Each reaction mixture was incubated at 37°C with 0.12 U of HIV-RT (Y188L) in the presence of 75 nM of each dNTP. The reactions were quenched after 40 min and the amount of tritium, incorporated by the polymerase, was measure on SPA beads by scintillation counting.

The first four reactions determined the amount of stalling versus elongation beyond the nonstandard base pair. [^{3}H]-TTP was used as the reporter nucleotide. Reaction #1 and #2 provided the positive control and background counts, respectively, #3 and #4 gave information about the success of primer extension beyond the pyDAD in the template. In the absence of dXTP (reaction #3), a faithful polymerase terminates the elongation at the position of pyDAD, resulting in only background counts in the SPA. Using the same experiment in the presence of dXTP (reaction #4), the enzyme should now go beyond the nonstandard base pair and incorporate radiolabelled TTP, leading to a signal in the SPA.

In the second set of reactions with [^{3}H]-dXTP as the radiolabelled reporter nucleotide, we directly monitored the incorporation of the radiolabelled nonstandard nucleoside triphosphate. Reaction #5 measured the total incorporation of radiolabelled dXTP while reaction #6 gave the background counts for the assay. Finally, reaction #7 monitored the amount of nonspecific dXTP incorporation opposite standard nucleotides in a template. After background subtraction, the ratio of #5 to #7 is an indicator of the specificity of the puADA incorporation by the polymerase.

Results from the SPA are presented in Figure 5-12. The positive control reaction #1 confirmed the polymerase activity of the HIV-RT preparation. At the same time, the reaction served as an "internal" standard. When compared to the counts from reaction #4, the amount of pausing at the nonstandard base pair can be estimated. Based on the counts in #4, only ~5% of the primer was extended to full-length product, indicating pausing of the polymerase, presumably at the pyDAD/puADA base pair. In addition, the data from reaction #3 are undistinguishable from the counts in #4, suggesting that at the present concentration of dXTP in the reaction mixture (75 nM), no real advantage is bestowed upon the enzyme to enable it to complete primer elongation.

The data from the [^{3}H]-dXTP experiments, however, indicate rapid and specific incorporation of the nonstandard nucleotide. In reaction #5, the counts are clearly above background (reaction #6), implying incorporation of the radiolabelled dXTP. To compare these data with the results from reactions #1-4, the almost 10-fold difference in specific activity of [^{3}H]-TTP and [^{3}H]-dXTP must be considered. The cpm were normalized by assuming that reaction #1 represented maximal performance under the experimental conditions. Based on the amount of incorporated [^{3}H]-TTP in reaction #1 and the ratio of specific activities of [^{3}H]-TTP (~129 Ci/mmol) and [^{3}H]-dXTP (~16 Ci/mmol), the primer extension experiments in the

presence of radiolabelled d(puADA)TP could yield a maximum signal of ~880 cpm. Considering the background noise of about 70 cpm, the detected 540 cpm for reaction #5 clearly indicates a significant amount of d(puADA)TP incorporation. The possibility of misincorporation of dXTP opposite standard nucleotides was checked in reaction #7. By replacing the pyDAD-containing template with a sequence-homologous standard

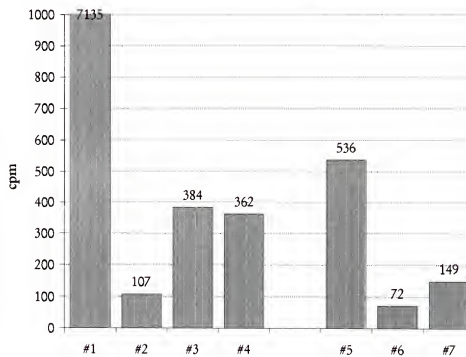


Figure 5-12. Graphic representation of primer extension reactions with pyDAD/puADA and HIV-RT (Y188L). Reactions #1-4 detect [^{3}H]-TTP incorporation while #5-7 use [^{3}H]-dXTP as a reporter nucleotide. Based on the positive control for polymerase activity (#1), the HIV-RT performs well with standard nucleotides. Reaction #2 and 6 are backgrounds for either radiolabelled nucleoside triphosphate. The SPA counts in reaction #3 and #4 represent a measurement for primer elongation beyond the nonstandard base pair. For reaction #5-7, the data suggest that HIV-RT (Y188L) incorporates [^{3}H]-dXTP opposite pyDAD in the template (reaction #5). At the same time, only minor unspecific incorporation of dXTP in the primer/template complex, made entirely from standard nucleotides, was observed (reaction #7). The actual counts per minute are plotted on top of each bar. The standard deviation for all measurements is $\pm$10%.

oligonucleotide, the amount of misincorporation of d(puADA)TP opposite one of the standard nucleotides in the presence of all the standard dNTPs was monitored. Based on the 150 cpm observed in reaction #7, we conclude that HIV-RT (Y188L) incorporates radiolabelled d(puADA)TP preferentially opposite pyDAD in the template.

We further conclude that, based on the SPA data for the incorporation of pyDAD and puADA by HIV-RT (Y188L), the enzyme rapidly incorporates the correct nonstandard nucleotide opposite its hydrogen bonding partner but then fails to continue primer elongation. The difference between these experiments and the previous primer extension and standing start experiments in Chapter 3 can be explained by the almost 3000-fold lower dNTP concentration in the SPA.

SPA with the pSBETa-9°N library - design

The polymerase activities of 48 colonies of BL21(DE3)pLysS-pSBETa-9°N, grown at 180-μL scale (microtiterplate) and 2-mL scale (Falcon tubes) were tested in primer extension experiments, followed by SPA analysis. Each heat-lysed sample was split into three aliquots which were then added to reaction mixture #1-#3 (Figure 5-13).

The three experiments are identical with reaction #1-4 in the preliminary tests with HIV-RT (Y188L). Reaction #1 was again used as the positive control to measure polymerase activity. In reaction #2 and 3, the difference of primer elongation beyond a pyDAD in the template in the presence (#2) or absence (#3) of dXTP was measured.

Model data for experimental analysis. To explain the data and a possible interpretation, as well as discuss the results from the SPA, the overall patterns that we expect to find are shown in a set of simulated data (Figure 5-14).

reaction mixtures

#1 → P5-S/T5-S1 + dATP + dGTP + TTP + dCTP + Polymerase

#2 → P5-S/T5-S2 + dATP + dGTP + TTP + dCTP + dXTP + Polymerase

#3 → P5-S/T5-S2 + dATP + dGTP + TTP + dCTP + Polymerase

P/T = 5'-biotin-GCG TAA TAC GAC TCA CTA TAG -3'
3'-GCG AGG ATG CTG AGT GAT ATC TGC TkG TAT - 5'

Figure 5-13. Primer extension and SPA experiments for the pSBETa-9°N library. T5-S1 in reaction #1 is the reference template with an adenosine in position k. T5-S2 in reactions #2 and #3 carries the pyDAD nonstandard nucleotide at position k. [^{3}H]-TTP was used in all three experiments as the reporter for successful primer elongation beyond the nonstandard base pair. Three experiments, using mixture #1 and water instead of the polymerase, were used to determine the background counts.

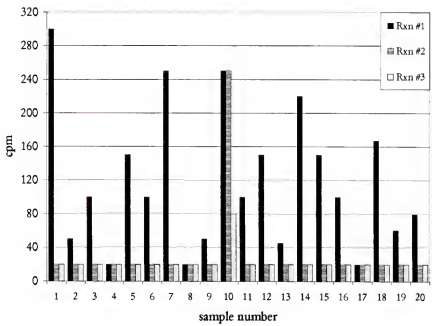


Figure 5-14. Ideal SPA results to illustrate the interpretation of the data. Some samples will lack any polymerase activity at all (samples #4,8, and 17). From the remaining preparations, >95% will show various levels of activity (reaction #1), due to differences in protein expression in the individual cultures. They, however, fail to incorporate puADA opposite pyDAD and therefore terminate primer elongation (reaction #2,3). Finally, a few samples will show the ability to extend the primer beyond the pyDAD in the template (sample #10).

Based on the model in Figure 5-14, we can make a basic distinction between samples with activity and samples without activity. The absence of activity can be explained by three possible scenarios: i) the random mutagenesis introduced a lethal mutation, ii) the protein expression failed in that particular sample, or iii) the expression vector is the product of self-ligation and therefore carries no insert in the first place. The different possibilities can be tested for by SDS-PAGE of the protein, restriction analysis of the plasmid, or simple regrowth of the colony for a second activity assay.

The majority of the samples are expected to express polymerase activity. Variations in the individual levels of activity can be attributed to differences in protein expression. Within this pool of enzymes, most will probably still behave like the wild-type enzyme, terminating the primer elongation in the presence of nonstandard bases. However, as indicated in sample #10 (Figure 5-14), a few candidates may show successful primer extension in the presence of d(puADA)TP (large cpm response for reaction #1 and #2), while in the absence of d(puADA)TP (reaction #3), the extension of the primer to full-length product is hindered, probably by the nonstandard nucleotide in the template. The plasmid DNA from these colonies will be isolated and taken into the next round of DNA shuffling.

SPA with the pSBETa-9°N library – experimental data

Reference experiments. In Figure 5-16, two of the samples were substituted for reference experiments with commercially available 9°N DNA polymerase. Experiment #1 and 42 were run in parallel and serve as identical control reactions. As expected for the wild-type enzyme, reaction #1 for both experiments resulted in incorporation of [^{3}H]-TTP. The lack of signal for either reaction #2 or #3 also demonstrate that the wild-type enzyme effectively terminates primer elongation in the presence of a nonstandard nucleotide in the template.

Cultures with no activity. Six colonies (#2, 18, 21, 23, 29, and 43) out of the 48 tested, roughly 13%, did not show polymerase activity in both expression experiments (Figure 5-15 and 5-16). In a restriction analysis of 10 of the 50 colonies prior to protein expression and activity screening, all plasmids carried the correct size insert-DNA, suggesting that the self-ligation of expression vector is not responsible for the lack of activity.

A lack of protein expression in these particular samples is also not very likely. First, the expression levels of the library have been very consistent in itself. Second, the chance that protein expression in the same sample fails in both experiments (Figure 5-15 and 5-16) is, based on our experience, very low. We therefore assume that the inactivity of the samples is actually the result of lethal mutagenesis.

Adopting this hypothesis, the data suggest that an average of 3-4 nucleotide changes in the 2.3 kb gene, which translates into 2-3 mutations on the protein level, knocks out ~13% of the enzyme population. Based on these numbers, a maximum of ~20 random amino acid changes are required to completely inactivate the enzyme. For the 770 amino acids in a DNA polymerase, this number indicates about 50 conserved amino acids (6.5% of the sequence), in which upon mutation of a single amino acid, results in complete inactivation of the enzyme. In contrast, structure and sequence studies suggest that only about 11 or 12 highly conserved amino acids (1.3% of sequence) are directly involved in DNA, metal, or dNTP binding (Blanco et al., 1991; Delarue et al., 1990). Disabling the catalytic site of the polymerase would therefore require a maximum of 64 random mutations.

The fivefold difference in the number of residues critical for enzyme function, as opposed to those involved in catalysis, underlines the importance of structural elements, possibly distant from the active site, for the overall performance of the enzyme. One might wonder, whether the thermostability of these enzymes could be represented in a fraction of

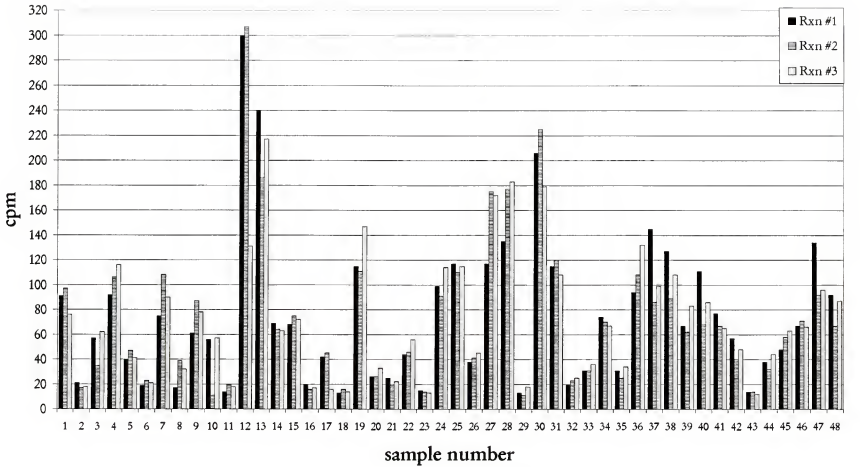


Figure 5-15. SPA data for cultures, grown in microtiterplates (180 μL cultures). Each enzyme preparation was tested in the three reaction mixtures (Figure 5-13) and the amount of radiolabelled TTP, incorporated by the polymerase, was determined by SPA. The background of the experiment is approximately 20 cpm as indicated.

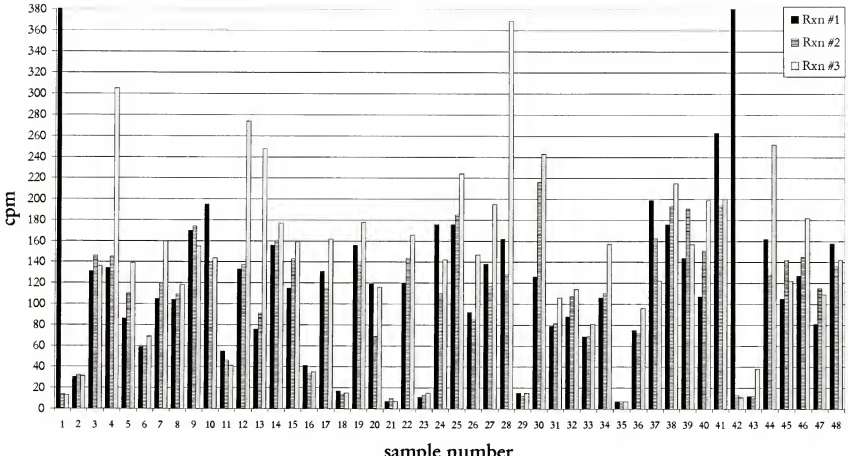


Figure 5-16. SPA data for cultures, grown in Falcon tubes (2-mL cultures). Each of the approximately 10-fold more concentrated enzyme samples (compared to Figure 5-15) was tested in reaction #1-#3 (Figure 5-13). Sample #1 and #42 are positive controls, using commercially available 9°N DNA polymerase (0.5 U/reaction). The background is approximately 20 cpm as indicated.

these additional conserved residues. The question becomes important when considering the purification procedure applied to our proteins. Any protein that carries a mutation that reduces its thermostability would be automatically selected out during heat lysis. In that case, our assay system would no longer screen only for polymerases able to incorporate pyDAD and puADA, but also for enzymes maintaining a high thermostability. Since structural flexibility of the polymerase could be one of the keys to better catalysts for nonstandard base pairs, the significance of random mutagenesis on thermostability should be studied in more details. Heat lysis of a series of protein expression cultures at various temperatures, followed by a simple activity assay, would quantify the fraction of proteins that are deselected based on their reduced thermostability rather than their lost catalytic function.

Cultures with activity. About 85% of the colonies expressed different levels of enzyme activity. This sample variation is clearly the result of the expected fluctuation in protein expression.

A good example for a successful candidate for consecutive rounds of shuffling is sample #12 in Figure 5-15. The results from the SPA indicate a similar elongation rate for reaction #1 and #2 plus an approximately twofold higher incorporation rate in reaction #2 than #3. Unfortunately, the result is not reproducible in the larger scale experiment.

In general, a data comparison between microtiterplate cultures (Figure 5-15) and Falcon tubes cultures (Figure 5-16) indicate some similarity in the overall pattern. For the individual samples, very similar cpm in reaction #1 through #3 were found.

This pattern was explained by the observation that the incorporation of [^{3}H]-TTP in these experiments is not exclusively the result of primer extension of our biotinylated primer-template complex. Ethidium bromide treatment and UV spectroscopy of the enzyme preparation after heat lysis suggest large amounts of DNA contamination. The UV reference

spectrum of 9°N DNA polymerase itself shows only a peak around 280 nm and no band at 260 nm, the UV spectrum of an aliquot of the polymerase sample has an absorption of approximately 1.25 A_{260} which is equivalent to 50-60 μg of double-stranded DNA in 20 μL of sample solution, typically used for the activity assay. Furthermore, the strong fluorescence, observed after addition of ethidium bromide proves the presence of double-stranded DNA in the sample.

We assume that the host-DNA is nicked and partly denatured during the heat treatment. During the following primer extension experiment, the polymerase not only elongates our biotinylated primer-template complex but also fills in the gaps in the damaged host-DNA. Upon transferring the labelled mixture into Flashplates or mixing with SPA beads, unspecific binding of the radiolabelled host-DNA to the plastic beads occurs, making an accurate quantification of [^{3}H]-TTP incorporation impossible. The additional background noise not only increases the vagueness of the entire assay but the side-reaction also effectively removes substrates from the reaction of interest, lowering its signal intensity. The combination of the two effects makes it impossible to draw any conclusions concerning the specificity of the polymerase for nonstandard nucleotides.

Conclusions

We have randomly mutated the gene of 9°N-7 DNA polymerase, taking advantage of natural error-frequency of *Taq* DNA polymerase. The desired mutation rate of 2-4 nucleotides per sequence was confirmed by DNA sequencing. The pool of gene sequences was then digested with DNase I, and the resulting 50-200 bp fragments were reassembled and PCR amplified successfully.

To screen the randomly mutagenized DNA polymerase for improved incorporation of nonstandard nucleotides, the PCR product was ligated into pSBETa, an expression vector that not only carries the T7-promoter region for efficient protein overexpression but also contains the gene for $tRNA^{AGA/AGG}$. Our data indicate that in the presence of the tRNA, the expression level of DNA polymerase in *E. coli* increases two to threefold.

The suitability of the primer extension-SPA screen to identify polymerases that incorporate nonstandard bases was tested with HIV-RT (Y188L). Although we detected the successful incorporation of d(puADA)TP with the help of radiolabelled material, the assay seems to amplify the effect of pausing, due to the very low dNTP concentrations used in the protocol. A straight-forward correction of this problem is to increase the dNTP concentration. Because the Flashplates are washed at least once prior to counting, the unincorporated material, which usually causes the concentration-dependent background counts, will be removed efficiently.

Over the course of these studies, it became very clear that one of the keys to success lies in the purification procedure applied to the overexpressed protein. The first problem encountered concerns the selection process during the isolation by heat lysis. To the best of our knowledge, this is the first attempt of directed evolution of a thermostable enzyme. Although several groups have more or less successfully increased the thermostability of enzymes from mesophilic ancestors using directed evolution, the effect of random mutagenesis on the thermostability of a protein is unknown. The current hypothesis concerning how enzymes achieve thermostability assumes that the observed heat resistance of an enzyme is the product of multiple ionic and hydrophobic interactions within the enzyme, contributing to the overall stability. Although the fraction of a protein actually involved in such interactions is

unknown, introducing random mutations in a thermophilic protein represents a potential risk to diminish its thermostability.

Applying the heat lysis procedure to isolate protein from the expression host, we may unintentionally end up selecting not only for new substrate specificity but also for maintenance of thermostability. On one hand, this selection could cause the loss of sequence diversity. The fraction of molecules that perform better with a nonstandard base pair but have additional mutations that limit their thermostability would be selected out during the protein isolation. On the other hand, this exact method could be applied to identify regions within a protein that must be conserved to maintain the thermal properties of the enzyme.

To address the question experimentally, the protein from a series of replica 96-well culture plates could be isolated by heat lysis of the host cells at various temperatures. An enzyme preparation showing no activity after heat lysis at 96°C might suddenly elongate a primer when isolated by heat-lysis at only 75°C. Performing such an experiment in a microtiterplate may allow some conclusions concerning the issue at a statistically significant level.

The second, more serious problem with our purification protocol is the large amount of DNA contamination in our protein purification procedure. The large amount of host-DNA not only explains the high magnesium concentration required in the SPA (20 mM, lower concentration showed a decreased enzyme activity, data not shown), but it also rationalizes the unexpectly high rates of primer extension in reaction #2 and #3 in the library screens (Figure 5-15 and 5-16).

To apply the primer extension-SPA methodology to the screening of our polymerase libraries, the DNA must be removed. A possible approach that does not require major modification of the purification protocol could be fractionated filtration through an ion-

exchange microtiterplate. In such a protocol, the cell pellet is resuspended in Tris-HCl buffer (pH 8) instead of the TN-buffer. After heat lysis, the material is then loaded onto an ion-exchange resin and could be eluted by raising the salt concentration in the washing buffer. The specific activity of the recovered material should be significantly higher and could again be tested for primer elongation.

The versatility of the approach has been shown by using the same protocol for two other hyperthermostable DNA polymerases. Gene and plasmid libraries of the exonuclease-deficient *Pfu* DNA polymerase and Vent (exo⁻) DNA polymerase were prepared. Together with 9°N DNA polymerase, family shuffling of the three enzymes may will result in some interesting new hybrids with novel properties.

Finally, the capacity of the assay has been tested on two occasions. The screening of up to 100 colonies by our primer extension/SPA-based screen has been demonstrated. In addition, in a second experiment, we processed 500 cultures in one day. At this point, the limits of available machinery were reached (PCR thermocycler, counter capacity). These 500 mutated enzymes, assayed in one day, represent more than twice the previous screening throughput in the Benner group for pyDAD and puADA.

APPENDIX A
GENETIC CODES AND AMINO ACIDS CODES

Genetic codes:

First base	Second base				Third base
	U=T	C	A	G	
U=T	Phe	Ser	Tyr	Cys	U=T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U=T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U=T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U=T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino Acids codes:

alanine	Ala	A
arginine	Arg	R
asparagine	Asn	N
aspartic acid	Asp	D
cysteine	Cys	C
glutamine	Gln	Q
glutamic acid	Glu	E
glycine	Gly	G
histidine	His	H
isoleucine	Ile	I
leucine	Leu	L
methionine	Met	M
phenylalanine	Phe	F
proline	Pro	P
serine	Ser	S
threonine	Thr	T
tryptophan	Trp	W
tyrosine	Tyr	Y
valine	Val	V

APPENDIX B
MATERIALS

B1 Chemicals

Ampicillin (sodium salt)	Sigma
Ammonium acetate	Fisher Sci.
Ammonium persulfate (APS)	Fisher Sci.
Acetic acid (glacial)	Fisher Sci.
Acetonitrile	Fisher Sci.
Boric Acid	Fisher Sci.
Bromophenol blue / Xylene cyanol (B/X-marker)	Sigma
Chloramphenicol	Sigma
Crystal violet	Fisher Sci.
Coomassie Brilliant Blue R250	BioRad
1,4-Dithiothreitol (DTT)	Sigma
EDTA (sodium salt)	Fisher Sci.
Ethanol (absolute)	Aaper Alc.Chem.Co.
Ethidium bromide	Fisher Sci.
Formamide	Fisher Sci.
Glycerol	Fisher Sci.
Guanidine hydrochloride	Sigma
Hydrochloric acid (36%)	Fisher Sci.
Isopropanol	Fisher Sci.
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Fisher Sci.
Kanamycin	Sigma
Magnesium acetate	Fisher Sci.
Magnesium sulfate	Fisher Sci.
Magnesium chloride (hexahydrate)	Fisher Sci.
Methanol	Fisher Sci.

PAGE solution 40 % (19:1) for nucleic acid	Fisher Sci.
PAGE solution 30 % (37.5:1) for protein	BioRad
Potassium chloride	Fisher Sci.
Sodium chloride	Fisher Sci.
Sodium acetate	Fisher Sci.
Sodium phosphate (mono/dibasic)	Fisher Sci.
Sodium hydroxide	Fisher Sci.
Sodium dodecylsulfate (SDS)	Fisher Sci.
Sodium sulfate (anhydrous)	Fisher Sci.
N,N,N,N,-Tetramethylethylendiamin (TEMED)	Fisher Sci.
Tris(hydroxymethyl)amino methane (base / HCl salt)	Fisher Sci.
Triton X-100	Sigma
Urea	Fisher Sci.

B2 Biochemicals / Molecular biology products

Agarose (ultra pure)	Life technologies
2'-Deoxy nucleoside triphosphates (dA, dG, dC, T) stock solution	Promega
FlashPlate® PLUS (streptavidin)	NEN Life Science
Oligonucleotides (standard)	IDT
Rapid DNA Ligation kit	Roche/Boehringer
Size marker: Lambda DNA – BstEII digest	New England Biolabs
Size marker: Lambda DNA – EcoR1/HindIII	New England Biolabs
Size marker: Oligo(dT)4-22 Ladder	Life Technologies
Size marker: SDS-PAGE standard – low range	BioRad
SPA - Random sequence reverse transcriptase enzyme assay	Amersham
TOPO XL PCR Cloning kit	Invitrogen
QIAEX II DNA purification system	Qiagen
QIAquick Gel extraction kit	Qiagen
Wizard Plus MiniPrep DNA purification system	Promega
Wizard DNA clean-up system	Promega

B3 Radioactive labeled material

Redivue™ [γ-^{32}P] adenosine-5'-triphosphate	Amersham/Pharmacia

[methyl,1',2'-^{3}H] thymidine 5'-triphosphate, ammonium salt	(> 5000 Ci/mmol) Amersham/Pharmacia
[methyl-^{3}H] thymidine 5'-triphosphate, ammonium salt	(90-130 Ci/mmol) Amersham/Pharmacia
[8-^{3}H] Deoxy adenosine-5'-triphosphate, tetrasodium salt	(40-60 Ci/mmol) ICN
[8-^{3}H] Deoxy guanosine-5'-triphosphate, tetrasodium salt	(5-25 Ci/mmol) ICN
[^{3}H] Tributyl tin hydride	(5-15 Ci/mmol) Moravek

B4 Enzymes

Restriction enzymes	New England Biolabs
Taq DNA polymerase	Promega
Pfu DNA polymerase	Stratagene
Pwo DNA polymerase	Roche/Boehringer
Vent DNA polymerase	New England Biolabs
Vent (exo-) DNA polymerase	New England Biolabs
DeepVent DNA polymerase	New England Biolabs
DeepVent (exo-) DNA polymerase	New England Biolabs
9°N DNA polymerase	New England Biolabs
Tth DNA polymerase	Promega
DNase I (RNase-free)	Roche/Boehringer
T4 Polynucleotide kinase	New England Biolabs
T4 DNA ligase	Promega
Phosphodiesterase I (Crotalus durissus terrificus)	Sigma and Roche/Boehringer
Alkaline Phosphatase (Bovine calf intestine)	Sigma and New England Biolabs

Enzymes were used with the buffers supplied by the manufacturer if not noted otherwise.

B5 Buffers and Media

Antibiotics stock solutions

ampicillin	50 mg/mL in water
kanamycin	10 mg/mL in water
chloramphenicol	34 mg/mL in ethanol abs.

Aqueous solutions were sterile filtered (0.22 μm Whatman) and stored at -20°C.

Buffers:

TBE-buffer (10x)	Tris base (108 g), Boric acid (55 g), EDTA (9.3 g) *ad* 1 Lt water
Tris-glycine buffer (10x)	Tris-HCl (30.2 g), Glycine (188 g), SDS (10%, 50 mL) *ad* 1 Lt water
HIV-RT buffer (3x)	Tris-HCl (500 µL, 300 mM, pH 7.2), $MgCl_2$ (15 µL, 1 M), KCl (300 µL, 1 M), DTT (3 µL, 1 M), EDTA (3 µL, 0.5 M, pH 8) *ad* 1 mL water
SPA Pol buffer (10x)	Tris-HCl (2 mL, 1 M, pH 8.8), Ammonium sulfate (1 mL, 1 M), Potassium chloride (333 µL,3 M) *ad* 10 mL water
TE buffer	Tris-HCl (5 mL, 200 mM, pH 7.5) EDTA (200 µL, 0.5 M, pH 8.0) *ad* 100 mL water
TN buffer	Tris-HCl (1 mL, 200 mM, pH 8.0) Sodium chloride (1 mL, 0.5 M) *ad* 10 mL water
Denature buffer	EDTA (40 µL, 0.5 M, pH 8.0) PAGE dye marker (20 µL), *ad* formamide (940 µL)
Crystal violet loading buffer (6x)	Crystal violet (50 µL, 2 mg/mL), EDTA (40 µL, 0.5 M, pH 8.0), Glycerol (300 µL) *ad* 1 mL water
Crush&Soak buffer	Ammonium acetate (9.64 g) Magnesium acetate (0.54 g) and SDS (0.25 g) *ad* 250 mL water
PAGE loading buffer	Bromophenol blue / Xylene cyanol mix (0.1 g) water (1 mL), and formamide (4 mL)

Media/Plates:

LB medium and LB/Agar medium were purchased from Fisher Scientific as dry powders. The corresponding solution was prepared following the manufacturer's instruction. The solutions were autoclaved at 121°C for 30 min. LB medium was stored at RT; LB/Agar plates were prepared by letting the sterilized medium cool to approximately 50°C (hand-warm). After addition of the particular antibiotics (ampicillin: 50 μg/mL; kanamycin: 30 μg/mL; chloramphenicol: 34 μg/mL) the solution was poured on sterile petri dishes and the agar was allowed to solidify. The plates were stored at 4°C.

B6 Bacterial strains and plasmids

HB 101	Clontech	F⁻ *hsd*S20 (r_B^- m_B^-) *sup*E44 *ara*14 *gal*K2 *lac*Y1 *pro*A *rps*L20 (Str^R) *xyl*-5 *mtl*-1 *rec*A13 *mcr*B *thi*-1 *leu*B6
TOP10	Invitrogen	F⁻ *mcr*BΔ (*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1 *nup*G
INVαF'	Invitrogen	F⁻ *end*A1 *rec*A1 *hsd*R17 (r_K^- m_K^+) *sup*E44 *thi*-1 *gyr*A96 *rel*A1 ϕ80*lac*ZΔM15 Δ (*lac*ZYA-*arg*F) U169
DH5α	Clontech	*deoR, endA1, gyrA96, hsdR17(r_K^- m_K^+), recA1, relA1, supE44, thi-1, Δ(lacZYA-argFV169), ϕ80δlacZΔM15, F, λ⁻*
BL21-Gold (DE3)pLysS	Stratagene	*E.coli* B F⁻ *ompT hsd*S(r_B^- m_B^-) *dcm*⁺ Tet^r *gal* (DE3) *end*A [pLysS Cam^r]
BL21(DE3) pLysS	Invitrogen	*E.coli* B F⁻ *ompT hsd*S(r_B^- m_B^-) *dcm*⁺ *gal* (DE3) [pLysS Cam^r]
BL21(DE3)	Novagen	*E.coli* B F⁻ *ompT hsd*S(r_B^- m_B^-) *dcm*⁺ Tet^r *gal* (DE3)

All strains except the BL21 family were purchased as electrocompetent cells. If required, more electrocompetent cells were prepared following the standard protocol (Ausubel et al., 1997).

pUC18	Pharmacia	
pBCSK+	Stratagene	
pCR®-XL TOPO	Invitrogen	
pET11a	Novagen	
pAll17	New England Biolabs	(Perler et al., 1992; Southworth et al., 1996)
pSBETa	H.H. Steinbiss	(Schenk et al., 1995)

B7 Equipment

Thermocycler	Perkin Elmer Genecycler 9700 (96 well format)
Vacuum concentrator	Eppendorf 5301 Vacuum Concentrator
Centrifuges	Eppendorf 5415 C Eppendorf 5417 R (refridgerated centrifuge)
Microtiterplate filtration unit	Whatman Polyfiltronics Univacs manifold VA02L w/ Unifilter plates (0.45 μm PVDF membrane)
UV/VIS Spectrophotometer	Varian Cary 1 Bio w/ sample changer and temp.control and Cary UV Win v.3.0
UV microtiter plate reader	Packard Instruments Spectra Count™ w/ PlateReader v.2.11
Gel Electrophoresis	BioRad Mini Sub Gel GT / Sub Gel GT system BioRad Mini Protean II system Gibco/Life technologies S2 / SA-32
Power supply	BioRad PowerPack 3000
Gel dryer	BioRad Gel dryer Model 583
Gel documentation	BioRad Gel Doc 1000 w/ Multianalyst v.1.0
Isotopic imaging	BioRad Molecular Imager GS-525 w/ Multianalyst v.1.0
Scintillation counting	Packard Instruments TopCount NTX Microplate Counter
Electroporator	BioRad E.coli pulser
Incubator	Fisher Sci. Isotemp.
Incubator Shaker	New Brunswick Sci.Co. Inc. Series 25
HPLC analytical	Waters Alliance System w/ Waters 996 PDA detector/Waters 600S Controller / Waters 616 Pump
HPLC preparative	Waters PrepLC 4000 w/ Waters 486 Abs.detector; - controlled by Waters Millenium^{32} ™ software

APPENDIX C
GENERAL METHODS IN MOLECULAR BIOLOGY

C1 Glycerol cultures:

In an Eppendorf tube, sterilized glycerol (100-150 μL) was mixed with an overnight culture in LB (800-900 μL) containing the appropriate antibiotics. The glycerol culture was immediately frozen in dry ice or liquid nitrogen and stored at -80°C.

C2 Overnight cultures:

LB-medium in a sterilized Erlenmeyer flask containing the appropriate antibiotics was inoculated with culture (from freezer stock, transformation, agar plates) using a sterile pipet tip (1-10 μL). The cells were grown for 12-16 hours in a shaker at a specific temperature.

C3 Isolation of plasmid DNA:

Plasmid DNA from cell cultures were isolated using the Wizard® Plus MiniPreps DNA purification system from Promega (Madison, WI). The protocol of the manufacturer was strictly followed.

C4 Desalting of double-strand DNA

Large DNA fragments (> 500 bp) were desalted using the Wizard® DNA clean-up system from Promega. For short DNA fragments (20-500 bp), the QIAEX II kit (Qiagen, Valencia CA) was used. In both cases, the protocols of the manufacturer were followed.

C5 Purification of double-strand DNA (preparative scale)

DNA fragments after restriction digestion or PCR products were purified by agarose gel electrophoresis. Agarose gels (0.7–1% in 0.5x TBE) using crystal violet (2 mg/mL stock solution; ~1 μL/mL gel) as a stain were poured. The samples were mixed with 6x crystal violet loading buffer (see Appendix 5B) and loaded on the agarose gel. The electrophoresis was run at a constant 80 V. Migration of the bands can be followed visually and the run was stopped when the fragments were sufficiently resolved.

The desired band was excised from the gel, using a fresh razor blade. The DNA was recovered from the gel piece using the QIAquick gel extraction kit from Qiagen. Elution with Tris-HCl (10 mM, pH 8) yielded DNA, which could directly be used for further enzymatic reactions.

C6 Determination of DNA concentration

The DNA concentration in samples was determined by UV absorbance spectroscopy. An aliquot from the DNA sample (1–5 μL) was added to a water- or buffer-filled cuvettes (1 mL volume, zero and baseline corrected) and the UV range from 360 to 200 nm was scanned. The absorbance difference at 260 nm was determined and the concentration calculated, based on:

$OD_{260} \times 100 : n = \text{nmoles}$ n = length of the oligonucleotide

and 1 OD_{260} = 33 μg ss-DNA
1 OD_{260} = 37 μg ds-DNA
1 OD_{260} = 50 μg plasmid DNA

C7 Restriction digestion of DNA

The DNA sample was mixed with the appropriate reaction buffer (supplied with the enzyme). For double digestions, the buffer recommendations from New England Biolabs were followed. A typical reaction mixture had a total volume of 40 μL and was incubated with 1-2 μL of enzyme preparation (~20-40 U) at 37°C in the thermocycler (minimizes evaporation) overnight.

C8 Phosphorylation of 5'-end of oligonucleotides

The method was used to attach an isotope-label on the 5'-end of oligodexoyribonucleotides. DNA (~0.03 OD analytical scale / 0.1 OD preparative scale) in water (7 μL), T4 reaction buffer (1 μL), and γ-^{32}P-ATP (1 μL analyt/ 2 μL prep.) was phosphorylated by addition of 10 U of T4-polynucleotide kinase (1 μL) and incubation at 37°C for 30 min. The enzyme was inactivated by heat treatment of the solution at 80°C for 10 min. Following addition of PAGE gel loading buffer, the analytical samples were ready for analysis. The preparative samples were precipitated with EDTA (0.2 μL; 0.5 M; pH 8.0), sodium acetate (1 μL; 3 M, pH 5.2), and 25 μL EtOH. After centrifugation, the supernatant was discharged and the pellet washed once with 50 μL ethanol (70%). The final pellet was dried in the vacuum concentrator and redissolved in water.

C9 Polyacrylamide gel electrophoresis (PAGE)

A typical 12% nucleic acid gels was prepared by mixing PAGE solution (40%, 30 mL), urea (42 g), TBE buffer (10x, 10 mL), and water (25 mL). Polymerization was induced by addition of APS (10%, 350 μL) and TEMED (50 μL) and the solution poured between two previously silylated glass plates. The gel solidifies within 1-2 hours and was stored overnight.

Previous to sample loading, the gel was preheated for 1 h at 75 W (~60°C). The electrophoresis was performed at 75 W, keeping the power constant. After the bromophenol marker had moved to the bottom, the current was stopped. Analytical gels transferred onto filter paper and Saran wrap and loaded in the gel dryer (80°C constant, 1 h). The gel was analyzed by autoradiology. Preparative gels were transferred onto Saran wrap and the product

bands identified by "UV shadowing". The bands were excised with a razor blade and the DNA isolated by crush & soak (Appendix C10).

C10 SDS-PAGE for protein analysis

Protein analysis by SDS-PAGE was performed with the Mini Protean II system (BioRad). For the resolving gel, a mixture of PAGE (30%, 4 mL), Tris-HCl (1.5 M, pH 8.8, 2.5 mL), SDS (10%, 100 μL), and water (3.3 mL) was polymerized by addition of APS (10%, 50 μL) and TEMED (5 μL). The solution was poured between the glass plates up to about 1 cm beneath the comb and covered with ethanol. After approximately 30 min, the ethanol was poured off and the stacking gel poured. The stacking solution contains PAGE (30%, 0.67 mL), Tris-HCl (0.5 M, pH 6.8, 1.26 mL), SDS (10%, 50 μL), water (3 mL), and polymerization was initiated with APS (10%, 25 μL) and TEMED (5 μL). The stacking gel takes about 20 min to completely polymerize. The amounts given are sufficient for two gels. The resulting 12% PAGE gel is used for separation of protein from 10 – 100 kDa.

After the electrophoresis is finished, the gels were removed from the glass plates and stained for 3-5 h in Coomassie blue solution (Sambrook & Manatis, 1983), followed by destaining in a mixture of water (50%), methanol (40%), and acetic acid (glacial, 10 %) overnight.

C11 DNA isolation from PAGE gels

The pieces of the excised PAGE gel were mixed with approximately 10 volume crush&soak buffer (Appendix 5B) in a Falcon tube. After incubating the tube at 37°C overnight in a shaker, the entire content was loaded in a syringe and filtered through a Whatman 0.2 μm sterile syringe filter. The DNA was recovered from the PAGE-free filtrate by ethanol precipitation. Small oligonucleotides (<20 bp) were desalted by dialysis against water.

C12 Oligonucleotide analysis by enzymatic digestion

Oligonucleotide (0.1 - 0.5 OD) in 2 mL water was mixed with 2 μL digest buffer (Tris-HCl, pH 8.3, 0.1 M; $MgCl_2$ 20 mM). The digestion was initiated by addition of phosphodiesterase (0.5 U) and alkaline phosphatase (5 U) and the sample incubated at 37°C for three to four hours. After quenching with TEAA (80 μL, 25 mM), the entire sample was analyzed by RP-HPLC (Nova-Pak® C18, 3.9x150 mm; A: TEAA (25 mM, pH 7), B: TEAA (25 mM, pH 7) with 20% acetonitrile; gradient from 0 - 20% B (10 min) 20 - 50% B (10 min)).

REFERENCES

Ambrose, B. J. B. & Pless, R. C. (1987). **DNA sequencing: Chemical methods.** *Methods Enzymol.* **152**, 522-538.

Angerer, B. & Ankenbauer, W. (1995). **Einbau ungewöhnlicher Nukleotide durch DNA-Polymerasen.** *Technischer Bericht, Boehringer Mannheim.*

Argoudelis, A. D. & Mizsak, S. A. (1976). **1-Methylpseudouridine, a metabolite of *Streptomyces platensis*.** *J. Antibiot.* **29**, 818-823.

Arnold, F. H. (1996). **Directed Evolution: Creating Biocatalysts for the Future.** *Chemical Engineering Science* **51**(23), 5091-5102.

Arnold, F. H. (1998). **Design by Directed Evolution.** *Acc. Chem. Res.* **31**, 125-131.

Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., Eds. (1997). **Short Protocols in Molecular Biology.** 3rd. edit. New York: John Wiley & Sons, Inc.

Bailly, C. & Waring, M. J. (1998). **The use of diaminopurine to investigate structural properties of nucleic acids and molecular recognition between ligands and DNA.** *Nucleic Acids Res.* **26**(19), 4309-4314.

Bakhanashvili, M., Avidan, O. & Hizi, A. (1996). **Mutational studies of human immunodeficiency virus type 1 reverse transcriptase: the involvement of residues 183 and 184 in the fidelity of DNA synthesis.** *FEBS Lett.* **391**, 257-262.

Barnes, W. M. (1992). **The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion.** *Protein Sci.*, 29-35.

Barton, D. H. R. & Subramanian, R. (1976). **Synthesis of deoxysugars and of deoxynucleosides from diol thiocarbonates.** *J. Chem. Soc. Chem. Commun.* **21**, 867-868.

Battersby, T. R., Ang, D. N., Burgstaller, P., Jurczyk, S., Bowser, M. T., Buchanan, D. D., Kennedy, R. T. & Benner, S. A. (1999). ***In vitro* selection of an Adenosine Receptor from a Library Incorporating a Cationic Nucleotide Analog.** *J. Am. Chem. Soc* in press.

Beaudry, A. A. & Joyce, G. F. (1992). **Directed evolution of an RNA enzyme.** *Science* **257**, 635-641.

Been, M. D. & Cech, T. R. (1988). **RNA as a RNA polymerase: net elongation of an RNA primer catalyzed bz the *Tetrahymena* ribozyme.** *Science* **239**, 1412-1416.

Bhattacharya, B. K., Devivar, R. V. & Revankar, G. R. (1995). **A Practical Synthesis of N1-Methyl-2'-Deoxy-psi-Uridine (psi-Tymidine) and its Incorporation into G-rich Triple Helix Forming Oligonucleotides.** *Nucleosides Nucleotides* **14**(6), 1269-1287.

Bjoerk, G. R. (1995). **Biosynthesis and Function of Modified Nucleosides.** In *tRNA: Structure, Biosynthesis, and Function* (Soell, D. & RajBhandary, U. L., eds.), pp. 165-205. ASM Press, Washington, D.C.

Blanco, L., Bernad, A., Blasco, M. A. & Salas, M. (1991). **A general structure for DNA-dependent DNA polymerases.** *Gene* **100**, 27-38.

Boosalis, M. S., Petruska, J. & Goodman, M. F. (1987). **DNA Polymerase Insertion Fidelity.** *J. Biol. Chem.* **262**(30), 14689-14696.

Boyer, P. L., Tantillo, C., Jacobo-Molina, A., Nanni, R. G., Ding, J., Arnold, E. & Hughes, S. H. (1994). **Sensitivity of wild-type human immunodeficiency virus type 1 reverse transcriptase to dideoxynucleotides depends on template length; sensitiivty of drug-resistant mutants does not.** *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4882-4886.

Breaker, R. R. (1997a). **DNA aptamers and DNA enzymes.** *Curr. Opin. Chem. Biol.* 1(1), 26-31.

Breaker, R. R. (1997b). **DNA enzymes.** *Nat. Biotechnol.* **15**, 427-431.

Breaker, R. R. & Joyce, G. F. (1994). **A DNA enzyme that cleaves RNA.** *Chem. Biol.* 1, 223-229.

Brinkmann, U., Mattes, R. E. & Buckel, P. (1989). **High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product.** *Gene* **85**, 109-114.

Bryant, F. R., Johnson, K. A. & Benkovic, S. J. (1983). **Elementary steps in the DNA polymerase I reaction pathway.** *Biochemistry* **22**(15), 3537-3546.

Cadwell, R. C. & Joyce, G. F. (1994). **Mutagenic PCR.** *PCR Methods Appl.* 3, S136-S140.

Carmi, N., Balkhi, S. R. & Breaker, R. R. (1998). **Cleaving DNA with DNA.** *Proc. Natl. Acad. Sci. U.S.A.* **95**, 2233-2237.

Carmi, N., Shultz, L. A. & Breaker, R. R. (1996). ***In vitro* selection of self-cleaving DNAs.** *Chem. Biol.* 3, 1039-1046.

Chartrand, P., Usman, N. & Cedergren, R. (1997). **Effect of structural modifications on the activity of the leadzyme.** *Biochemistry* **36**, 3145-3150.

Chen, G. F. & Inouye, M. (1990). Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes. *Nucleic Acids Res.* **18**(6), 1465-1473.

Chen, G. T. & Inouye, M. (1994). Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* **8**, 2641-2652.

Chu, C. K., Reichmann, U., Watanabe, K. A. & Fox, J. J. (1977). Nucleosides. 104. Synthesis of 4-Amino-5-(D-ribofuranosyl)pyrimidine C-nucleosides from 2-(2,3-O-Isopropylidene -5-O-trityl-D-ribofuranosyl)acetonitrile. *J. Org. Chem.* **42**, 711-714.

Cohen, K. A., Hopkins, J., Ingraham, R. H., Pargellis, C., Wu, J. C., D.E.H., P., Kinkade, P., Warren, T. C., Rogers, S., Adams, J., Farina, P. R. & Grob, P. M. (1991). Characterization of the Binding Site for Nevirapine (BI-RG-587), a Nonnucleoside Inhibitor of Human Immunodeficiency Virus Type-1 Reverse Transcriptase. *J. Biol. Chem.* **266**(22), 14670-14674.

Cole, J. L. (1996). Approaches to High-Volume Screening Assays of Viral Polymerases and Related Proteins. *Methods Enzymol.* **275**, 310-328.

Coll, M., Wang, A. H. J., van der Marel, G. A., van Boom, J. H. & Rich, A. (1986). Crystal Structure of a Z-DNA Fragment Containing Thymidine/2-Aminoadenosine Base Pairs. *J. Biomol. Struct. Dyn.* **4**(2), 157-172.

Conner, B. N., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. (1982). The molecular structure of d(CpCpGpG), a fragment of right-handed double helical A-DNA. *Nature (London)* **295**, 294-299.

Cook, A. F., Vuocolo, E. & Brakel, C. L. (1988). Synthesis and hybridization of a series of biotinylated oligonucleotides. *Nucleic Acids Res.* **16**, 4077-4095.

Crameri, A., Raillard, S. A., Bermudez, E. & Stemmer, W. P. C. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature (London)* **391**, 288-291.

Crameri, A., Whitehorn, E. A., Tate, E. & Stemmer, W. P. C. (1996). Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling. *Nat. Biotechnol.* **14**, 315-319.

Creighton, S., Bloom, L. B. & Goodman, M. F. (1995). Gel Fidelity Assay Measuring Nucleotide Misinsertion, Exonucleolytic Proofreading, and Lesion Bypass Efficiencies. *Methods Enzymol.* **262**, 232-256.

Crick, F. H. C. (1968). The origin of the genetic code. *J. Mol. Biol.* **38**, 367-379.

Cubellis, M. V., Rozzo, C., Montecucchi, P. & Rossi, M. (1990). Isolation and sequencing of a new b-galactosidase-encoding archaebacterial gene. *Gene* **94**, 89-94.

Cuenoud, B. & Szostak, J. W. (1995). A DNA metalloenzyme with DNA ligase activity. *Nature (London)* **375**, 611-614.

Dabrowski, S. & Kur, J. (1998). Cloning and Expression in *Escherichia coli* of the Recombinant His-Tagged DNA Polymerases from *Pyrococcus furiosus* and *Pyrococcus woesei*. *Proteins Exp. Purif.* **14**, 131-138.

Dai, X., De Mesmaeker, A. & Joyce, G. F. (1995). Cleavage of an amide bond by a ribozyme. *Science* **267**, 237-240.

Das, K., Ding, J., Hsiou, Y., Clark Jr, A. D., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A. J., Boyer, P. L., Clark, P., Smith Jr, R. H., Kroeger-Smith, M. B., Michejda, C. J., Hughes, S. H. & Arnold, E. (1996). Crystal structure of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181 Cys HIV-1 RT drug-resistant mutant. *J. Mol. Biol.* **264**, 1085-1100.

Davies, D. B. (1978). Conformations of nucleosides and nucleotides. *Progress in NMR Spectroscopy* **12**, 135-225.

Davis, D. R. (1995). Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**(24), 5020-5026.

Debyser, Z., De Vreese, K., Dnops-Gerrits, P. P., Baekelandt, V., Bhikhabhai, R., Strandberg, B., Pauwels, R., Anne, J., Desmyter, J. & De Clercq, E. (1993). Kinetics of Different Human Immunodeficiency Virus Type 1 Reverse Transcriptase Resistant to Human Immunodeficiency Virus Type 1-Specific Reverse Transcriptase Inhibitors. *Mol. Pharmacol.* **43**, 521-526.

Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Eng.* **3**(6), 461-467.

Derbyshire, V., Pinsonneault, J. K. & Joyce, C. M. (1995). Structure-Function Analysis of 3'-5' Exonuclease of DNA Polymerases. *Methods Enzymol.* **262**, 363-385.

Desai, U. J. & Pfaffle, P. K. (1995). Single-Step Purification of a Thermostable DNA Polymerase Expressed in *Escherichia coli*. *BioTechniques* **19**, 780-784.

Dewey, T. M., Zyzniewski, C. & Eaton, B. E. (1996). The RNA world: functionalized diversity in a nucleotide by carboxyamidation of uridine. *Nucleosides Nucleotides* **15**, 1611-1617.

Diederichsen, U. (1998). Selectivity of DNA Replication: The Importance of Base-Pair Geometry over Hydrogen Bonding. *Angew. Chem., Int. Ed. Engl.* **37**(12), 1655-1657.

Drew, H. R. & Dickerson, R. E. (1981). Structure of a B-DNA Dodecamer. *J. Mol. Biol.* **151**, 535-556.

Ellington, A. D. & Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature (London)* **346**, 818-822.

Engelke, D. R., Krikos, A., Bruck, M. E. & Ginsburg, D. (1990). Purification of *Thermus aquaticus* DNA Polymerase Expressed in *Escherichia coli*. *Anal. Biochem.* **191**, 396-400.

Eritja, R., Horowitz, D. M., Walker, P. A., Ziehler-Martin, J. P., Boosalis, M. S., Goodman, M. F., Itakura, K. & Kaplan, B. E. (1986). Synthesis and properties of oligonucleotides containing 2'-deoxynebularine and 2'-deoxyxanthosine. *Nucleic Acids Res.* **14**(20), 8135-8153.

Evans, T. A. & Seddon, K. R. (1997). Hydrogen bonding in DNA - a return to the *status quo*. *J. Chem. Soc. Chem. Commun.*, 2023-2024.

Fan, N., Rank, K. B., Evans, D. B., Thomas, R. C., Tarpley, W. G. & Sharma, S. K. (1995). Simultaneous mutations at Tyr-181 and Tyr-188 in HIV-1 reverse transcriptase prevents inhibition of RNA-dependent DNA polymerase activity by the bisheteroarylpiperazine (BHAP) U-90152s. *FEBS Lett.* 370, 59-62.

Farr, R. N., Outten, R. A., Cheng, J. C. & Daves, G. D. (1990). C-glycoside synthesis by palladium-catalyzed iodoaglycon-glycal coupling. *Organometallics* **9**, 3151-3156.

Faulhammer, D. & Famulok, M. (1996). The Ca^{2+} ion as a cofactor for a novel RNA-cleaving deoxyribozyme. *Angew. Chem., Int. Ed. Engl.* 35(23/24), 2837-2841.

Fiala, G. & Stetter, K. O. (1986). *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100deg.C. *Arch. Microbiol.* **145**, 56-61.

Florian, J. & Leszczynski, J. (1995). Theoretical Investigation of the Molecular Structure of the pk DNA Base Pair. *J. Biomol. Struct. Dyn.* **12**(5), 1055-1062.

Geyer, C. R. & Sen, D. (1997). Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enyzme. *Chem. Biol.* **4**, 579-593.

Giver, L., Gershenson, A., Freskgard, P. O. & Arnold, F. H. (1998). Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12809-12813.

Goodman, M. F., Creighton, S., Bloom, L. B. & Petruska, J. (1993). Biochemical Basis of DNA Replication Fidelity. *Crit. Rev. Biochem. Mol. Biol.* **28**(2), 83-126.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. (1983). The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell* 35, 849-857.

Hager, A. J. & Szostak, J. W. (1997). Isolation of novel ribozymes that ligate AMP-activated RNA substrates. *Chem. Biol.* **4**, 607-617.

Hampel, A. & Cowan, J. A. (1997). A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chem. Biol.* **4**, 513-517.

Hasson, M. S., Schlichting, I., Moulai, J., Taylor, K., Barrett, W., Kenyon, G. L., Babbitt, P. C., Gerlt, J. A., Petsko, G. A. & Ringe, D. (1998). Evolution of an enzyme active site: The structure of a new crystal from of muconate lactonizing enzyme compared with mandelate racemase and enolase. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10396-10401.

Horlacher, J. Enzymatic Incorporation of Non-Standard Base Pairs into DNA. Ph.D. Thesis, ETH Nr. 11084, Swiss Federal Institute of Technology, Zurich, Switzerland, 1995.

Horlacher, J., Hottinger, M., Podust, V. N., Hübscher, U. & Benner, S. A. (1995). Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with nonstandard hydrogen bonding pattern. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6329-6333.

Howard, F. B., Chen, C., Cohen, J. S. & Miles, H. T. (1984). Poly(d2NH$_2$A-dT): Effect of 2-amino substituent on the B to Z transition. *Biochem. Biophys. Res. Commun.* **118**(3), 848-853.

Hsu, M., Inouye, P., Rezende, L., Richard, N., Li, Z., Prasad, V. R. & Wainberg, M. A. (1997). Higher fidelity of RNA-dependent DNA mispair extension by M184V drug-resistant than wild-type reverse transcriptase of human immunodeficiency virus type 1. *Nucleic Acids Res.* **25**(22), 4532-4536.

Hu, X. Y., Shi, Q. W., Yang, T. & Jackowski, G. (1996). Specific replacement of consecutive AGG codons results in high-level expression of human cardiac troponin T in Escherichia coli. *Proteins Exp. Purif.* **7**(3), 289-293.

Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998). Structure of a Covalently Trapped Catalytic Complex of HIV-1 Reverse Transcriptase: Implications for Drug Resistance. *Science* **282**, 1669-1675.

Illangaskare, M. & Yarus, M. (1997). Small molecule - substrate interactions with a self-aminoacylating ribozyme. *J. Mol. Biol.* **268**(3), 631-639.

Innis, M. A., Myambo, K. B., Gelfand, D. H. & Brow, M. A. D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9436-9440.

James, K. D. & Ellington, A. D. (1997). Surprising fidelity of template-directed chemical ligation of oligonucleotides. *Chem. Biol.* **4**, 595-606.

Jannasch, J. W., Wirsen, C. O., Molyneaux, S. J. & Langworthy, T. A. (1992). Comparative Physiological Studies on Hyperthermophilic Archaea isolated from Deep-Sea Hot Vents with emphasis on *Pyrococcus* strain GB-D. *Appl. Environ. Microbiol.* **58**, 3472-3481.

Jeong, W. & Shin, H. C. (1998). Supply of the argU gene product allows high-level expression of recombinant human interferon-a2a in *Escherichia coli*. *Biotechnol. Lett.* 20(1), 19-22.

Jestin, J. L., Kristensen, P. & Winter, G. (1999). A Method for the Selection of Catalytic Activity Using Phage Display and Proximity Coupling. *Angew. Chem., Int. Ed. Engl.* 38(8), 1124-1127.

Joyce, C. M. & Steitz, T. A. (1995). Minireview - Polymerase structures and functions: Variations on a theme? *J. Bacteriol.* 177(22), 6321-6329.

Joyce, G. F. (1989). Amplification, mutation and selection of catalytic RNA. *Gene* 82, 83-87.

Joyce, G. F. (1994). In vitro evolution of nucleic acids. *Curr. Opin. Struct. Biol.* 4, 331-336.

Joyce, G. F. (1998). Nucleic acid enzymes: Playing with a fuller deck. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5845-5847.

Kahl, J. D. & Greenberg, M. M. (1999). Introducing Structural Diversity in Oligonucleotides via Photolabile, Convertible C5-Substituted Nucleotides. *J. Am. Chem. Soc* 121(4), 597-604.

Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. (1992). Mechanism and Fidelity of HIV Reverse Transcriptase. *J. Biol. Chem.* 267(36), 25988-25997.

Kawase, Y., Iwai, S., Inoue, H., Miura, K. & Ohtsuka, E. (1986). Studies on nucleic acid interactions I. Stabilities of mini-duplexes ($dG_2A_4XA_4G_4$ - $dC_2T_4YT_4C_2$) and self-complementary d(GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. *Nucleic Acids Res.* 14(19), 7727-7736.

Kawase, Y., Iwai, S. & Ohtsuka, E. (1989). Synthesis and Thermal Stability of Dodecadeoxyribonucleotides Containing Deoxyinosines Pairing with Four Major Bases. *Chem. Pharm. Bull.* 37(3), 599-601.

Kellam, P., Boucher, C. A. B. & Larder, B. A. (1992). Fifth mutation in human immunodeficiency virus type 1 reverse transcriptase contributes to the development of high-level resistance to zidovudine. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1934-1938.

Kerr, S. G. & Anderson, K. S. (1997). Pre-Steady-State Kinetic Characterization of Wild Type and 3'-Azido-3'-deoxythymidine (AZT) Resistant Human Immunodeficiency Virus Type 1 Reverse Transcriptase: Implication of RNA Directed DNA Polymerization in the Mechanism of AZT Resistance. *Biochemistry* 36, 14064-14070.

Kim, B. & Loeb, L. A. (1995). Human immunodeficiency virus reverse transcriptase substitutes for DNA polymerase I in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 684-688.

Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J. & Rich, A. (1973). Three-dimensional structure of yeast phenylalanine transfer RNA: Folding of the polynucleotide chain. *Science* 179, 285-288.

Knutsen, L. J. S., Judkins, B. D., Newton, R. F., Scopes, D. I. C. & Klinkert, G. (1985). Synthesis of imidazo-fused bridgehead-nitrogen 2'-deoxyribo-C-nucleosides: Coupling-elimination reactions of 2,5-anhydro-3,4,6-tri-O-benzoyl-D-allonic acid. *J. Chem. Soc., Perkin Trans. 1* 3, 621-630.

Kodra, J. T. Chemistry and Enzymology of an Expanded Genetic Alphabet. Ph.D. Thesis, ETH Nr. 12916, Swiss Federal Institute of Technology, Zurich, Switzerland, 1998.

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). Crystal structure at 3.5Å Resolution of HIV-1 Reverse Transcriptase Complexed with an Inhibitor. *Science* 256, 1783-1790.

Kong, H., Kucera, R. B. & Jack, W. E. (1993). Characterization of a DNA polymerase from the Hyperthermophilie Archaea *Thermococcus litoralis*. *J. Biol. Chem.* 268(3), 1965-1975.

Kopp, E. B., Miglietta, J. J., Shrutkowski, A. G., Shih, C. K., Grob, P. M. & Skoog, M. T. (1991). *Nucleic Acids Res.* 19, 3035-3039.

Kuchta, R. D. (1996). Isotopic Assays of Viral Polymerases and Related Proteins. *Methods Enzymol.* 275, 241-257.

Kurger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E. & Cech, T. R. (1982). Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* 31, 147-157.

Larder, B. A., Darby, G. & Richman, D. D. (1989). HIV with Reduced Sensitivity to Zidovudine (AZT) Isolated During Prolonged Therapy. *Science* 243, 1731-1734.

Larder, B. A., Kellam, P. & Kemp, S. D. (1991). Zidovudine resistance perdicted by direct detection of mutations in DNA from HIV-infected lymphocytes. *AIDS* 5(2), 137-144.

Larder, B. A. & Kemp, S. D. (1989). Multiple Mutations in HIV-1 Reverse Transcriptase Confer High-Level Resistance to Zidovudine (AZT). *Science* 246, 1155-1158.

Larder, B. A., Kemp, S. D. & Harrigan, P. R. (1995). Potential Mechanism for substained Antiretroviral Efficancy of AZT-3TC Combination Therapy. *Science* 269, 696-699.

Latham, J. A., Johnson, R. & Toole, J. J. (1994). The application of a modified nucleotide in aptamer selection: novel thrombin aptamers containing 5-(1-pentynyl)-2'-deoxyuridine. *Nucleic Acids Res.* 22(14), 2817-2822.

Leach, A. R. & Kollman, P. A. (1992). Theoretical Investigations of Novel Nucleic Acid Bases. *J. Am. Chem. Soc* 114, 3675-3683.

Li, Y. & Breaker, R. R. (1999). Phosphorylating DNA with DNA. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2746-2751.

Li, Y. & Sen, D. (1996). A catalytic DNA for porphyrin metallation. *Nat. Struct. Biol.* 3, 743-747.

Limbach, P. A., Crain, P. F. & McCloskey, J. A. (1994). The Modified Nucleosides of Rna - Summary. *Nucleic Acids Res.* 22(12), 2183-2196.

Liu, D. R., Magliery, T. J., Pastrnak, M. & Schultz, P. G. (1997). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 10092-10097.

Lohse, P. A. & Szostak, J. W. (1996). Ribozyme-catalysed amino-acid transfer reactions. *Nature (London)* 381, 442-444.

Lorimer, I. A. J. & Pastan, I. (1995). Random recombination of antibody single chain Fv sequences after fragmentation with DNase I in the presence of Mn^{2+}. *Nucleic Acids Res.* 23(15), 3067-3068.

Lorsch, J. R. & Szostak, J. W. (1994). *In vitro* selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* 33, 973-982.

Lu, C. & Erickson, H. P. (1997). Expression in *Escherichia coli* of the Thermostable DNA Polymerase from *Pyrococcus furiosus*. *Proteins Exp. Purif.* 11, 179-184.

Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A. & Mathur, E. J. (1991). High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* 108, 1-6.

Lutz, M. J. Erkennung von Nucleinsaeuren-Analoga durch Polymerasen. Ph.D. Thesis, ETH Nr. 12254, Swiss Federal Institute of Technology, Zurich, Switzerland, 1997.

Lutz, M. J., Held, H. A., Hottinger, M., Hübscher, U. & Benner, S. A. (1996). Differential discrimination of DNA polymerases for variants of the non-standard nucleobase pair between xanthosine and 2,4-diaminopyrimidine, two components of an expanded genetic alphabet. *Nucleic Acids Res.* 24, 1308-1313.

Lutz, M. J., Horlacher, J. & Benner, S. A. (1998). Recognition of a Non-Standard Base Pair by thermostable DNA Polymerases. *Bioorg. Med. Chem. Lett.* 8, 1149-1152.

Lutz, S. & Benner, S. A. (1999). Simple One-Pot Synthesis of a 2'-Tritium Labeled C-Deoxynucleoside. *Bioorg. Med. Chem. Lett.* 9, 723-726.

Marky, L. A., Blumenfeld, K. S., Kozlowski, S. & Breslauer, K. J. (1983). Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d(CGCGAATTCGCG): Evidence for hairpin formation. *Biopolymers* **22**, 1247-1257.

Marx, A., Spichty, M., Amacker, M., Schwitter, U., Huebscher, U., Bickle, T. A., Maga, G. & Giese, B. (1999). Probing interactions between HIV-1 reverse transcriptase and its DNA substrate with backbone-modified nucleotides. *Chemistry & Biology* **6**, 111-116.

Mathur, E. J. (1996). Purified thermostable pyrococcus furiosus DNA polymerase I. Stratagene (La Jolla, CA), United States.

Maxam, A. M. & Gilbert, W. (1980). Sequencing End-labeled DNA with base-specific chemical cleavage. *Methods Enzymol.* **65**, 499-560.

Moore, J. C. & Arnold, F. H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. *Nat. Biotechnol.* **14**, 458-467.

Moran, S., Ren, R. X. F. & Kool, E. T. (1997a). A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. U.S.A.* 94, 10506-10511.

Moran, S., Ren, R. X. F., Rumney, S. & Kool, E. T. (1997b). Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc* **119**, 2056-2057.

Muramatsu, T., Nishikawa, K., Nemoto, F., Kuchino, Y., Nishimura, S., Miyazawa, T. & Yokoyama, S. (1988a). Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature (London)* **336**, 179-181.

Muramatsu, T., Yokoyama, S., Horie, N., Matsuda, A., Ueda, T., Yamaizumi, Z., Kuchino, Y., Nishimura, S. & Miyazawa, T. (1988b). A Novel Lysine-substituted Nucleoside in the First Position of the Anticodon of Minor Isoleucine tRNA from *Escherichia coli*. *J. Biol. Chem.* **263**(19), 9261-9267.

Niehaus, F., Frey, B. & Antranikian, G. (1997). Cloning and characterization of a thermostable a-DNA polymerase from the hyperthermophilic archaeon Thermococcus sp. TY. *Gene* 204, 153-158.

O'Leary, D. J. & Kishi, Y. (1994). Preferred Conformation of C-Glycosides. 13. A Comparison of the Conformation Behavior of Several C-, N-, and O-Furanosides. *J. Org. Chem.* **59**, 6629-6636.

Orgel, L. E. (1968). Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381-393.

Pandey, V. N., Kaushik, N., Rege, N., Sarafianos, S. G., Yadav, P. N. S. & Modak, M. J. (1996). Role of Methionine 184 of Human Immunodeficiency Virus Type-1 Reverse

Transcriptase in the Polymerase Function and Fidelity of DNA Synthesis. *Biochemistry* 35, 2168-2179.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Broka, C., Rice, J. A., Itakura, K. & Breslauer, K. J. (1982). Premelting and melting transitions in the d(CGCGAATTCGCG) self-complementary duplex in solution. *Biochemistry* 21, 428-436.

Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase (beta), a DNA template-primer, and ddCTP. *Science* 264, 1891-1903.

Peracchi, A., Beigelman, L., Scott, E. C., Uhlenbeck, O. C. & Herschlag, D. (1997). Involvement of a specific metal ion in the transition of the Hammerhead Ribozyme to its catalytic conformation. *J. Biol. Chem.* 272(43), 26822-26826.

Perler, F. B., Comb, D. G., Jack, W. E., Moran, L. S., Qiang, B., Kucera, R. B., Benner, J., Slatko, B. E., Nwankwo, D. O., Hempstead, S. K., Carlow, C. K. S. & Jannasch, H. (1992). Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5577-5581.

Perler, F. B., Kumar, S. & Kong, H. (1996). Thermostable DNA Polymerases. *Advances in Protein Chemistry* 48, 377-435.

Perrin, D. D. & Armarego, W. L. F. (1988). *Purification of Laboratory Chemicals*. 3 rd. edit, Butterworth-Heinemann Ltd., Oxford, UK.

Piccirilli, J. A. The Evolution of Biological Catalysis. Ph.D. Thesis, Harvard University, Cambridge, MA, 1989.

Piccirilli, J. A., Krauch, T., MacPherson, L. J. & Benner, S. A. (1991a). A direct route to 3-(D-ribofuranosyl)pyridine nucleosides. *Helv. Chim. Acta* 74, 397-406.

Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. (1990). Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature (London)* 343, 33-37.

Piccirilli, J. A., Moroney, S. E. & Benner, S. A. (1991b). A C-nucleotide base pair: Methylpseudouridine-directed incorporation of formycin triphosphate into RNA catalysed by T7 RNA polymerase. *Biochemistry* 30, 10350-10356.

Pluthero, F. G. (1993). Rapid purification of high-activity *Taq* DNA polymerase. *Nucleic Acids Res.* 21(20), 4850-4851.

Proba, K., Wörn, A., Honegger, A. & Plückthun, A. (1998). Antibody scFv Fragments without Disulfide Bonds Made by Molecular Evolution. *J. Mol. Biol.* 275, 245-253.

Prudent, J. R., Staunton, J. & Schultz, P. G. (1995). Probing the structural determinants of a catalytic RNA with isomerase activity. *J. Am. Chem. Soc* **117**, 10145-10146.

Reardon, J. E. & Miller, W. H. (1990). Human Immunodeficiency Virus Reverse Transcriptase. *J. Biol. Chem.* **265**(33), 20302-20307.

Ren, J., Esnouf, R. M., Hopkins, A. L., Jones, E. Y., Kirby, I., Keeling, J., Ross, C. K., Larder, B. A., Stuart, D. I. & Stammers, D. K. (1998). 3'-Azido-3'-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9518-9523.

Rich, A. (1962). On the problems of evolution and biochemical information transfer. In *Horizons in Biochemistry* (Kasha, M. & Pullmann, B., eds.), pp. 103-126. Academic Press, New York.

Robertson, D. L. & Joyce, G. F. (1990). Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature (London)* **344**, 467-468.

Robertson, M. P. & Miller, S. L. (1995). Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. *Science* **268**, 702-705.

Rosenberg, A. H., Goldman, E., Dunn, J. J., Studier, F. W. & Zubay, G. (1993). Effects of Consecutive AGG Codons on Translation in Escherichia coli, Demonstrated with a Versatile Codon Test System. *J. Bacteriol.* **175**(3), 716-722.

Roth, A. & Breaker, R. R. (1998). An amino acid as a cofactor for a catalytic polynucleotide. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6027-6031.

Roy, K. B. & Miles, H. T. (1983). Tautomerism and ionization of xanthosine. *Nucleosides Nucleotides* **2**(3), 231-242.

Sabeti, P. C., Unrau, P. J. & Bartel, D. P. (1997). Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.* **4**(10), 767-774.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*. 1 ed. Springer Advanced Texts in Chemistry (Cantor, C. R., Ed.), Springer Verlag, New York.

Sakthivel, K. & Barbas, C. F. (1998). Expanding the Potential of DNA for Binding and Catalysis: Highly functionalized dUTP Derivatives that are Substrates for Thermostable DNA Polymerases. *Angew. Chem., Int. Ed. Engl.* **37**(20), 2872-2875.

Sambrook, J., Fritsch, E. F. & Maniatis, T. Sequencing by the Maxam-Gilbert Method. In *Molecular Cloning: A laboratory manual*. 2 ed, Cold Spring Harbor Laboratory Press, New York, 1989, 13.78-13.101.

Sambrook, J., Fritsch, E. F. & Maniatis, T., Eds. *Molecular Cloning: A laboratory manual*. 2 ed, Cold Spring Harbor Laboratory Press, New York, 1989.

Sardana, V. V., Emini, E. A., Gotlib, L., Graham, D. J., Lineberger, D. W., Long, W. J., Schlabach, A. J., Wolfgang, J. A. & Condra, J. H. (1992). Functional Analysis of HIV-1 Reverse Transcriptase Amino Acids Involved in Resistance to Multiple Nonnucleoside Inhibitors. *J. Biol. Chem.* **267**(25), 17526-17530.

Schenk, P. M., Baumann, S., Mattes, R. & Steinbiss, H. H. (1995). Improved High-Level Expression System for Eukaryotic Genes in *Escherichia coli* Using T7 RNA Polymerase and Rare Arg-tRNAs. *BioTechniques* **19**(2), 196-200.

Seelig, B. & Jaeschke, A. (1999). A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167-176.

Seyhan, A. A., Amaral, J. & Burke, J. M. (1998). Intracellular RNA cleavage by the hairpin ribozyme. *Nucleic Acids Res.* **26**(15), 3494-3504.

Singh, J., Wise, D. S. & Townsend, L. B. (1991). 2'-Deoxy-3',5'-O-(tetraisopropylidisiloxanyl) pseudouridine. In *Nucleic acid chemistry* (Townsend, L. B. & Tipson, R. S., eds.), Vol. 4, pp. 96-98. 4 vols. Wiley and Son, New York.

Sioud, M. (1999). Application of preformed hammerhead ribozymes in the gene therapy of cancer (Review). *Int. J. Mol. Med.* **3**(4), 381-384.

Skandalis, A., Encell, L. P. & Loeb, L. A. (1997). Creating novel enzymes by applied molecular evolution. *Chem. Biol.* **4**, 889-898.

Smerdon, S. J., Jäger, J., Wang, J., Kohlstaedt, L. A., Chirino, A. J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1994). Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3911-3915.

Southworth, M. W., Kong, H., Kucera, R. B., Ware, J., Jannasch, H. W. & Perler, F. B. (1996). Cloning of thermostable DNA polymerases from hyperthermophilic marine archaea with emphasis on *Thermococcus* sp.9°N-7 and mutations affecting 3'-5' exonuclease activity. *Proc. Natl. Acad. Sci. U.S.A.* **93**(May), 5281-5285.

Spence, R. A., Kati, W. M., Anderson, K. S. & Johnson, K. A. (1995). Mechanism of Inhibition of HIV-1 Reverse Transcriptase by Nonnucleoside Inhibitors. *Science* **267**, 988-993.

Stahlhut, M. W. & Olsen, D. B. (1996). Expression and Purfication of Retroviral HIV-1 Reverse Transcriptase. *Methods Enzymol.* **275**, 122-132.

Steitz, T. A. (1987). The Klenow Fragment Structure Suggests Mechanisms for Fidelity and Processivity of DNA Polymerase I. In *Biological Organization: Macromolecular*

Interactions at High Resolution (Brunett, R. M. & Vogel, H. J., eds.), pp. 45-55. Academic Press, New York.

Steitz, T. A. (1998). A mechanism for all polymerases. *Nature (London)* 391, 231-232.

Steitz, T. A. & Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* 90, 6498-6502.

Stemmer, W. P. C. (1994a). DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91, 10747-10751.

Stemmer, W. P. C. (1994b). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature (London)* 370, 389-391.

Strazewski, P. & Tamm, C. (1990). Replication experiments with nucleotide base analogues. *Angew. Chem., Int. Ed. Engl.* 29, 36-57.

Studier, F. W. & Moffatt, B. A. (1986). Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes. *J. Mol. Biol.* 189, 113-130.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Duebendorff, J. W. (1990). Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.* 185, 60-89.

Suzuki, M., Baskin, D., Hood, L. & Loeb, L. A. (1996). Random mutagenesis of *Thermus aquaticus* DNA polymerase I: concordance of immutable sites *in vivo* with the crystal structure. *Proc. Natl. Acad. Sci. U.S.A.* 93(9670-9675).

Sweasy, J. B. & Loeb, L. A. (1992). Mammalian DNA polymerase b can substitute for DNA Polymerase I during DNA Replication in *Escherichia coli*. *J. Biol. Chem.* 267(3), 1407-1410.

Switzer, C., Moroney, S. & Benner, S. A. (1989). Enzymatic incorporation of a new base pair into DNA and RNA. *J. Am. Chem. Soc* 111, 8322-8323.

Switzer, C. Y., Moroney, S. E. & Benner, S. A. (1993). Enzymatic Recognition of the Base Pair between Isocytidine and Isoguanosine. *Biochemistry* 32, 10489-10496.

Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. & Imanaka, T. (1997). Characterization fo DNA Polymerase from *Pyrococcus* sp. Strain KOD1 and Its Application to PCR. *Appl. Environ. Microbiol.* 63(11), 4504-4510.

Tantillo, C., Ding, J., Jacobo-Molina, A., Nanni, R. G., Boyer, P. L., Hughes, S. H., Pauwels, R., Andries, K., Janssen, P. A. J. & Arnold, E. (1994). Locations of anti-AIDS drug binding sites and resistance mutations in the three-dimensional structure of HIV-1 reverse transcriptase. *J. Mol. Biol.* 243, 369-387.

Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. (1997). RNA-catalysed carbon-carbon bond formation. *Nature (London)* 389, 54-57.

Tomita, K., Ueda, T. & Watanabe, K. (1999). The presence of pseudouridine in the anticodon alters the genetic code: a possible mechanism for assignment of the AAA lysine codon as asparagine in echinoderm mitochondria. *Nucleic Acids Res.* 27(7), 1683-1689.

Tor, Y. & Dervan, P. B. (1993). Site-specific enzymatic incorporation of an unnatural base, *N6*-(6-aminohexyl)isoguanosine, into RNA. *J. Am. Chem. Soc* 115, 4461-4467.

Townsend, L. B., Wise, D. S. & Earl, R. A. (1978). 2,4-Dichloro-5-(b-D-ribofuranosyl) pyrimidines and substituted derivatives, pp. 8. University of Utah, Salt Lake City, Utah, US.

Tsubouchi, A. & Bruice, T. C. (1995). Phosphonate ester hydrolysis catalysed by two lanthanum ions. Intramolecular nucleophilic attack of corrdinated hydroxide and Lewis acid activation. *J. Am. Chem. Soc* 117(28), 7399-7411.

Tuerk, C. & Gold, L. (1990). Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science* 249, 505-510.

Uemori, T., Ishino, Y., Toh, H., Asada, K. & Kato, I. (1993). Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.* 21(2), 259-265.

Ueno, T. & Mitsuya, H. (1997). Comparative enzymatic study of HIV-1 reverse transcriptase resistant to 2',3'-dideoxynucleotide analogs using the single-nucleotide incorporation assay. *Biochemistry* 36(5), 1092-1099.

Unrau, P. J. & Bartel, D. P. (1998). RNA-catalysed nucleotide synthesis. *Nature (London)* 395, 260-263.

Van Aerschot, A., Mag, M., Herdewijn, P. & Vanderhaeghe, H. (1989). Double protection of the heterocyclic base of xanthosine and 2'-deoxyxanthosine. *Nucleosides Nucleotides* 8, 159-178.

Vesnaver, G. & Breslauer, K. J. (1991). The contribution of DNA single-stranded order to the thermodynamics of duplex formation. *Proc. Natl. Acad. Sci. U.S.A.* 88, 3569-3573.

Voegel, J. J. & Benner, S. A. (1994). Nonstandard hydrogen bonding in duplex oligonucleotides. The base pair between an acceptor-donor-donor pyrimidine analog and a donor-acceptor-acceptor purin analog. *J. Am. Chem. Soc* 116, 6929-6930.

von Krosigk, U. & Benner, S. A. (1995). A pH-dependent triple helix formation by an oligonucleotide containing a pyrazine donor-donor-acceptor base. *J. Am. Chem. Soc* 117, 5361-5362.

Wade, K., Wade, Y., Ishibashi, F., Gojobori, T. & Ikemura, T. (1992). Codon usage tabulated from the genebank genetic sequence data. *Nucleic Acids Res.* 20, 2111-2118.

Watson, J. D. & Crick, F. H. C. (1953). **Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid.** *Nature (London)* **171**, 964-967.

Werner, C., Krebs, B., Keith, G. & Dirheimer, G. (1976). **Specific cleavages of pure tRNAs by plumbous ions.** *Biochim. Biophys. Acta* **432**, 161-175.

Wiegand, T. W., Janssen, R. C. & Eaton, B. E. (1997). **Selection of RNA amide synthases.** *Chem. Biol.* **4**, 675-683.

Wilson, C. & Szostak, J. W. (1995). ***In vitro* evolution of a self-alkylating ribozyme.** *Nature (London)* **374**, 777-782.

Wilson, J. E., Aulabaugh, A., Caligan, B., McPherson, S., Wakefield, J. K., Jablonski, S., Morrow, C. D., Reardon, J. E. & Furman, P. A. (1996). **Human Immunodeficiency Virus Type-1 Reverse Transcriptase - contribution of Met-184 to binding of nucleotide 5'-triphosphate.** *J. Biol. Chem.* **271**(23), 13656-13662.

Wise, D. S., Earl, R. A. & Townsend, L. B. (1978). ***The chemistry and biological activity of C-nucleosides related to pseudouridine.*** Chemistry and biology of nucleosides and nucleotides (Harmon, R. E., Robins, R. K. & Townsend, L. B., Eds.), Academic Press, New York.

Witkin, E. M. & Roegner-Maniscalco, V. (1992). **Overproduction of DnaE Protein (a-Subunit of DNA polymerase III) Restores Viability in a Conditional Inviable *Escherichia coli* Strain Deficient in DNA Polymerase I.** *J. Bacteriol.* **174**(12), 4166-4168.

Woese, C. (1967). **The evolution of the genetic code.** In *The genetic code*, pp. 179-195. Harper & Row, New York.

Yamao, F., Andachi, Y., Muto, A., Ikemura, T. & Osawa, S. (1991). **Levels of tRNAs in bacterial cells as affected by amino acid usage in proteins.** *Nucleic Acids Res.* **19**(22), 6119-6122.

Zhang, B. & Cech, T. R. (1997). **Peptide bond formation by *in vitro* selected ribozymes.** *Nature (London)* **390**, 96-100.

Zhang, J., Dawes, G. & Stemmer, W. P. C. (1997). **Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening.** *Proc. Natl. Acad. Sci. U.S.A.* **94**(9), 4504-4509.

Zhao, H. & Arnold, F. H. (1997a). **Functional and nonfunctional mutations distinguished by random recombination of homologous genes.** *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7997-8000.

Zhao, H. & Arnold, F. H. (1997b). **Optimization of DNA shuffling for high fidelity recombination.** *Nucleic Acids Res.* **25**(6), 1307-1308.

Zhao, H. & Arnold, F. H. (1999). **Directed evolution converts subtilisin E into a functional equivalent of thermitase.** *Protein Eng.* **12**(1), 47-53.

BIOGRAPHICAL SKETCH

On June 17, 1969, I was born in St.Gallen (Switzerland) to Ernst and Gudrun Lutz-Oehler. From 1976 to 1985, I attended primary and secondary school in Flawil (Switzerland). After moving to Basel (Switzerland), I completed in three-year organic synthetic vocational training program at Ciba-Geigy. In 1988, I joined the Biotechnology Research Division at Ciba-Geigy as a laboratory assistant, working on the isolation, characterization, and derivatization of secondary metabolites.

From 1989 to 1992, I studied chemistry and chemical engineering at the Winterthur University of Applied Sciences (Winterthur, Switzerland). For my B.Sc. thesis on the stereoselective synthesis of D-amino acids using metallo-organic chemistry, I worked under the supervision of Dr. U. Michel. After graduating, I participated in a two-year program, leading to a M.Sc. in biotechnology. After the first year at the Winterthur University of Applied Sciences (Winterthur, Switzerland), I moved to Middlesbrough (UK) for the second year of studies and to complete my Master's thesis at the University of Teesside (Middlesbrough, UK). In the laboratory of Dr. G. Singh, I worked on the biosynthesis of substituted clavulanic acid by *Streptomyces clavuligerus*.

After graduating with a M.Sc. in biotechnology in 1995, I joined the research group of Dr. S. A. Benner at the University of Florida (Gainesville, FL).

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Steven A. Benner, Chairman
Professor of Chemistry

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Tomas Hudlicky
Professor of Chemistry

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David E. Richardson
Professor of Chemistry

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Jon D. Stewart
Assistant Professor of Chemistry

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Ben M. Dunn
Distinguished Professor of Biochemistry and
Molecular Biology

This dissertation was submitted to the Graduate Faculty of the Department of Chemistry in the College of Liberal Arts and Sciences and the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 1999

Dean, Graduate School